AUDITORY INTERACTION IN TWO-TONE OCTAVE COMPLEXES

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE
GENEESKUNDE
AAN DE ERASMUS UNIVERSITEIT TE ROTTERDAM
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF. DR. B. LEIJNSE
EN VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN.
DE OPENBARE VERDEDIGING ZAL PLAATS VINDEN OP
WOENSDAG 29 JUNI 1977 DES NAMIDDAGS
TE 4.15 UUR PRECIES

DOOR

PIETER JAN JOHANNES LAMORÉ

GEBOREN TE NIEUW BEIJERLAND

PROMOTOR : PROF. DR. G. VAN DEN BRINK
CO-REFERENTEN: PROF. DR. E. H. HUIZING
PROF. DR. IR. R. PLOMP

mijn vader en moeder

---

aan Emmy, Agnite, Willemijn en Aldert

VOORWOORD

Velen ben ik zeer veel dank verschuldigd. Het verschijnen van dit proefschrift geeft mij de gelegenheid allen die tot het tot stand komen ervan een meer of minder directe bijdrage geleverd hebben met name te noemen.

Allereerst noem ik mijn ouders, die mijn studie en werk steeds met zo grote belangstelling hebben gevolgd. Met vele anderen bedroeft het mij dat zij dit verschijnen niet meemaken.

Vervolgens noem ik mijn promotor, Prof. Dr. G. van den Brink. Gert, ik heb in de loop van de jaren veel vrijheid opgeëist, zowel bij de keuze van het onderwerp van mijn onderzoek als bij het gaan van de wegen. Ik heb het je daarbij niet gemakkelijk gemaakt. Dat je me bent blijven accepteren en corrigeren stemt me tot dankbaarheid.

Prof. Dr. E.H. Huizing dank ik ervoor dat hij, bij zijn vele verkzaamheden, als coreferent tijd vrijmaakte om het manuscript te lezen en adequaat te bekritizeren.

Een weer geheel andere inbreng is afkomstig van de tweede coreferent Prof. Dr. Ir. R. Plomp. Reinier, je hebt een vrij permanente bemoeienis gehad met het onderzoek, o.a. omdat het onderwerp geënt is op het onderwerp van een hoofdstuk uit je proefschrift. Een ding van wat me bij deze bemoeienis is opgevallen wil ik expliciet noemen. Je verstaat als weinig anderen de kunst om iemand vanuit zijn eigen positie te stimuleren en te corrigeren.

Binnen de vakgroep "Natuurkunde en Technologie" noem ik met name Dr. M. Rodenburg en Drs. J. Verschuure. Marius, jouw bijdrage is geweest die van de kritische hoorder met veel gevoel voor de kern van de zaak. Vooral de laatste jaren hebben wij vaak lange gesprekken gehad over het onderzoek en over de publicaties. Hartelijk dank daarvoor. Hans, jou dank ik voor je enthousiasme in onze discussies. Ik vermeld ook dat jij de pulsatiedrempelmetingen sterk gestimuleerd hebt.

Voor wat betreft de wetenschappelijke bijdragen van

buiten de Erasmus Universiteit vermeld ik discussies en contacten met Dr. Ir. F.A. Bilsen, Dr. J.H. ten Kate en Dr. Ir. T.J.F. Buunen van de werkgroep Biologische Natuurkunde (afdeling Technische Natuurkunde) van de T.H. Delft en met Dr. G.F. Smoorenburg en Dr. Ir. T. Houtgast van het Instituut voor Zintuigfysiologie R.V.O.-T.N.O. te Soesterberg. Afzonderlijk dank ik Ir. W. Schoonderbeek die in het kader van zijn kandidaats- en afstudeeronderzoek nauw betrokken is geweest bij het onderzoek.

Technische bijdragen binnen het kader van de vakgroep werden geleverd door M.P. Brocaar, A.J.J. Maas. C. Verweij, C.J. Keemink, H. Roggeveen, C. Goedegebuur en Hans van der Bijll (practicant). Michael, Ronald, Kees, Kees, Hans en Cor, jullie leverden bijdragen in de ontwikkeling van de apperatuur en jullie waren bereid - al of niet over de muren van researchgroepen heen - allerlei werken en werkjes uit te voeren. Ik dank jullie voor jullie interesse in het onderzoek, waarbij ik met name Michael noem voor zijn bijdrage in de ontwikkeling van het model. E.E.E. Frietman van de groep Biologische Natuurkunde (afdeling Technische Natuurkunde) van de T.H. Delft leende mij voor enige tijd de Allen Avionics vertragingslijn uit.

Het proefschrift werd op juist gebruik van de engelse taal gecorrigeerd door Dr. R.H. Bathgate te Eindhoven. Hij heeft niet alleen met veel zorg het engels gecorrigeerd, maar zich ook in het onderwerp verdiept en met grote kennis van zaken de tekst becommentarieerd op betoogtrant. Dr. Bathgate, ik denk met genoegen terug aan de inspirerende commentaren en de discussies.

De figuren in dit proefschrift werden voor een deel met grote nauwgezetheid en geduld getekend en gefotografeerd door J. Melchior en voor het overige deel getekend en gefotografeerd door de medewerkers van de Audiovisuele Dienst.

Dit proefschrift is niet denkbaar zonder de inzet van de proefpersònen. Het zijn, in volgorde van komst, Ir. A.P. Eland, W. Joosten, Drs. M.E. Eland, G. Punt, A. van der Spek, H. Jonker, Ir. J.H. Meijer, Drs. J. Verschuure en J.N. Kroon.

Prof. D.C. den Haan dank ik voor zijn bereidheid mij, met opstelling en geluidarme box, een tijdlang gastvrijheid te verlenen in het pand Westzeedijk 140.

Het Ministerie van Onderwijs en Wetenschappen dank ik voor de financiele bijdrage in de vertaalkosten van dit proefschrift.

Ir. P.A.M. Maas en Mevr. L.M. Terhorst-ten Wolde van de

Deftse Universitaire Pers dank ik voor hun adviezen inzake de vormgeving van dit proefschrift.

De lezer moet niet gering denken over de hoeveelheid typewerk die aan het verschijnen van dit boek is voorafgegaan. Een groot deel van de eerste versie werd getypt door Mej. T. van Reijn. Twee volledige versies en nog meer werden getypt door Mevr. E.C. Lamoré-van Veen, weliswaar niet bij nacht en ontij maar toch dicht in die buurt. Many thanks to Mrs. J.S. Hemsley who typed the final manuscript fast and accurately. Hand en spandiensten werden daarbij verleend door Mej. M.G. van Kruining en Mevr. N.J. Dijkdrenth-Duijzend. Ik vermeld nog dat het plaatje op de omslag een computerplot is, verkregen m.b.v. een programma geschreven door Drs. J. Verschuure.

Ik begon dit voorbericht met mijn ouders, ik eindig vol met diegenen die mij het naast staan: Emmy, Agnite, Willemijn en Aldert. Zij zijn een bron van vreugde en inspiratie geweest zowel in goede als in kwade dagen. Behalve aan mijn ouders is daarom dit proefschrift aan hen opgedragen.

o	a	p	w	g	e	d	r
a	g	n	i	t	e	r	a
e	n	a	l	d	e	r	t
w	i	l	l	e	m	y	n
d	t	d	e	m	m	y	g
a	e	e	m	m	y	m	e
v	e	r	y	d	e	o	n
n	y	t	n	m	n	a	a

CONTENTS

LIST OF SYMBOLS
USED IN THE DESCRIPTION OF THE EXPERIMENTAL RESULTS

2-AFC	:	two-alternative forced choice procedure.
Δ	:	time between the end of the masker pulse and the onset of the test-tone pulse in forward masking (see Fig. 3.1).
f_1	:	frequency of the lower component in two-tone stimuli and frequency of the masker in the pure-tone masking and pulsation-threshold experiments.
L_1	:	level of the lower-frequency component and level of the masker. The levels are expressed as decibels sound pressure level (dB SPL) *re* 2.10^{-4} μb, or relative to the absolute threshold of the corresponding stimulus (dB SL).
f_2	:	frequency of the higher component in the two-tone stimuli.
f_t	:	frequency of the test tone in simultaneous masking, forward masking, pulsation-threshold and onset-threshold experiments.
L_t	:	level of the test tone.
f_i	:	frequency of the (third) interfering tone in some of the pitch experiments with octave complexes.
L_i	:	level of this interfering tone.
f_m	:	frequency of the stimulus in the experiments in which a comparison is made between forward masking, onset threshold, and pulsation threshold.
L_m	:	level of this stimulus tone.
ϕ_1	:	phase of the lower-frequency component as used in the presentation of the data. See Chapter 3.
ϕ_{2a}	:	phase of the higher-frequency component in the complex waveform at the entrance to the external auditory meatus. The relation between ϕ_{2a} and ϕ_1 is given in Fig. 3.3.

ϕ_2 : phase of the higher-frequency component in the complex waveform of the movement of the basilar membrane at the stapes. A positive value of the signal corresponds to a deflection towards the scala vastibuli.

$\phi_{max}(\phi_{min})$: phase ϕ_1 at which the loudness of the higher-frequency component is maximum (minimum).

$f_{tmax}(f_{tmin})$: maximum (minimum) value of f_t at which the phase has any effect on the pulsation threshold for octave complexes.

LIST OF SYMBOLS
USED IN THE DESCRIPTION OF THE MODEL

V_1 : input voltage for the lower-frequency component (f_1).

V_2 : input voltage for the higher-frequency component (f_2).

V_{uc} : output voltage for the two-tone signals.

$V_{u1}(V_{u2})$: output voltage for component I (II).

V_p : output voltage corresponding to the plateau in the dynamic characteristic for component I.

$\phi_2(\phi_1)$: phase of the higher- (lower-) frequency component in the input signal of the model.

$\phi_{2max}(\phi_{2min})$: phase of component II for which the output for the complex is maximum (minimum).

$\Delta\phi_2$: difference in phase of component II at the two entrances of the model.

τ : delay of the output signal of the second block with respect to that of the first one.

f_p : $\frac{1}{\tau}$

CHAPTER 1

GENERAL INTRODUCTION

1.1. Subject of the present investigations

It is well known that two simultaneously presented pure tones give rise to a sensation known as "beats" when their frequency ratio is slightly different from m:n, where m and n are small integers; the combinations 400/802 Hz and 400/603 Hz are examples of tone pairs which will give rise to beats. The beats are perceived as a periodic fluctuation in loudness, timbre, pitch or combinations of these. Although the phenomenon of beats has been known for a long time, relatively few quantitative data on their perceptual characteristics are available in the literature.

Stimuli consisting of two components with a frequency ratio slightly different from m:n can be considered as being exactly harmonic, while one component has a continuously running phase. The use of components with a frequency ratio of exactly m:n makes it possible to study the perceptual phenomena for particular phase relations separately. Stimuli with particular phase relations between the components are sometimes called "frozen beats". The present investigation is focussed on the interaction in two-tone stimuli in which the tones have a frequency ratio 1:2, which will be called "two-tone octave complexes" from now on. One of the main aims of the investigation is to quantify the perceptual changes due to variation of the phase.

The above mentioned two-tone stimuli represent a special class of stimuli. Many investigations of the interaction between pure tones have been carried out for frequency differences less than about a third of an octave (which correspond to the "critical band"). However, the frequency differences in stimuli consisting of two tones with a frequency ratio of m:n are usually larger than the critical bandwidth.

According to the present state of our knowledge, there are three main ways in which stimuli are processed in the auditory system. The first of these is frequency analysis. This may be illustrated as follows. A pure-tone stimulus gives rise to a travelling wave on the basilar membrane in the inner ear. The envelope of this wave has a roughly triangular form, sloping steeply at the apical end and shallowly at the basal end. This envelope is called the "mechanical stimulation pattern". The hair cells along the basilar membrane are connected to single nerve fibres. At each place on the membrane, there is a certain relation between the amplitude of the mechanical wave and the number of spikes per second in the neuron. As a consequence a pattern called the "activity pattern" or "excitation pattern" is generated in the neural system. Now the frequency analysis implies that a given position on the basilar membrane and a given nerve fibre have maximum sensitivity for stimulation at a certain frequency. Although the frequency selective action of the auditory system is well established, a lack of knowledge still exists as to the relation between the mechanical frequency selectivity on the basilar membrane and the frequency selectivity found in single neural units.

The second main mode of processing is time-structure detection. Many electrophysiological investigations have revealed that, up to a certain frequency, the moments at which action potentials in a single neural unit occur are related to the temporal structure of the stimulus.

The third main mode of processing is that of lateral suppression or contrast enhancement. A stimulus, e.g. a pure tone, presented to the ear has a certain masking effectiveness which can be considered as a measure of its "internal strength". The addition of a second pure tone to the stimulus can lead to a decrease in this masking effectiveness. Although more energy is applied, the perceived "strength" of the first tone decreases; the tone is "suppressed". This phenomenon is observed when the test tone is presented non-simultaneously with the masker, e.g. during the remaining state of adaptation after the masker is switched off. The "suppression" also manifests itself as a decrease in the loudness of a tone due to the presence of another one. This effect is also responsible for contrast enhancement in a stimulus with various frequency components. The phenomenon has been extensively studied by Houtgast (1972, 1973, 1974a and 1974b).

Frequency analysis and time-structure detection have been used in the literature to explain the phase effects observed in two-tone octave complexes. One explanation relates these effects to interference of a second harmonic of the lower-frequency component, generated in the ear due to distortion, with the higher-frequency component actually presented. The other main explanation describes the phase effects as due to variations in the temporal pattern of the stimulus. No general agreement as to the relative merits of these two explanations has been reached in the literature.

The second main aim of the present investigations is to look for new arguments for or against the above mentioned explanations. The lateral-suppression effect will be introduced as a new element in the discussion. The subject of investigation is discussed further in section 2.5, after the survey of the relevant literature.

1.2. Main lines of auditory research

Two main methods in auditory research are psychophysics and electrophysiology. In psychophysics data are collected by people who listen to selected sounds and assess them. This approach has the advantage that the auditory system is regarded as a whole. However, it often has the drawback that one cannot be certain which part of the system is specifically involved in a given phenomenon. Conclusions on this point can only be drawn by making assumptions and testing them.

In electrophysiological experiments data are collected from one element in the auditory system (often in an experimental animal), e.g. a neuron or a group of neurons; the function of the measured properties in the whole system is often unknown. An example is the timing of the signals that pass along the neural pathway. In general both methods are needed to gain a complete insight into the operation of the auditory system.

The present study is based on psychophysical measurements. In general, sounds of a given duration have three perceptual attributes. First, they can be located on a tonal scale: one stimulus may sound higher than another one. This attribute is called "pitch". The second characteristic is a location on a "loudness scale". Thirdly, two sounds which have the same pitch and loudness, may still have a different "timbre". It is rather difficult to define an absolute measure of these sensations. The attributes "pitch" and "loudness" are quantified usually by a matching procedure, in which the stimulus is presented in alternation with a pure tone called the "test tone" or "matching tone". The pitch of the stimulus is then expressed as the frequency of the matching tone which gives the same pitch as that of the stimulus. Similarly, the loudness of the stimulus is expressed as the level of the matching tone which gives the same loudness as the stimulus. The procedure becomes more difficult when the pitch or the loudness of a component of a complex signal has to be matched. The question of how to quantify timbre is more difficult to answer, since timbre is a multidimensional quantity. However, the measurement of timbre is not involved in our investigations.

Another way of quantifying the output of the auditory system is by means of a "threshold procedure", as used e.g. in masking experiments. Here a relatively high-level pure tone (the "masker") and a second tone called the "maskee" or "test tone" are presented simultaneously, and the level of the test tone is adjusted until this tone is just inaudible. The level obtained is called the "masked threshold of detectability". Here the notion "threshold" implies the detection of the boundary between audibility and inaudibility.

It was assumed for a long time that the results of masking experiments reflected some kind of internal excitation in the auditory system. Close comparison of electrophysiological and psychophysical data over the past decade has activated the development of a new psychophysical method, the "pulsation-threshold method" (Houtgast, 1972), which gives results agreeing closely with those of single-cell recordings in the peripherical auditory system. The pulsation-threshold method is the best psychophysical procedure for finding some kind of internal representation of the auditory stimuli known so far. It has, therefore, been widely used in the present investigations.

The pulsation-threshold method is a "threshold procedure" too, though in this case the threshold is one between continuity and pulsation, not between audibility and inaudibility as in masking experiments. The position of a threshold is always determined according to a certain criterion, on the basis of which the observer decides under which conditions the test tone is on one side of the boundary and under which conditions on the other. A problem with threshold procedures is that the results for different observers may show systematic differences because different criteria are used. This sometimes also holds for the same experiment performed by the same observer on different days. Although the criterion problem is an important one, it rarely has very serious consequences for the reproducibility of the results.

1.3. Survey of the contents of this thesis

Chapter 2 contains a review of the published experimental data on and explanations of the phase effects in the field of two-tone stimuli, followed by a more precise definition of the scope of the present investigations. The methods of measurement are discussed in Chapter 3. Chapters 4 and 5 present the results of our experiments on the sensations in two-tone octave complexes, in particular on the pitch and the loudness of the higher-frequency component as functions of level and phase for different frequency combinations. Chapters 6-9 are devoted to the considerations of how the stimuli are represented "internally" under the various phase and level conditions. This involves discussion of the current explanations of phase effects. In Chapter 10 we consider the pulsation-threshold method itself, to decide whether more than one dynamic system is involved in the results. Finally, Chapter 11 presents a model which may describe the phase effects.

1.4. Terminology

The lower-frequency component of the two-tone stimuli is called component I, its frequency f_1 and its level L_1. The higher-frequency component is called component II, its frequency f_2 and its level L_2. The levels are expressed as decibels sound pressure level (dB SPL) *re* $2.10^{-4}\mu b$. When the level of a stimulus is given relative to the absolute threshold of that stimulus, the expression "sensation level"

(SL) is used.

When the masked threshold of component II is determined, component I being the masker, we speak of pure-tone masking and call component II the test tone. Generally in the pure-tone masking experiments and in the pulsation-threshold experiments the level and the frequency of the masker are called L_1 and f_1, respectively, and the level and the frequency of the test tone f_t and L_t, respectively. The value of f_t is not necessarily equal to $2f_1$.

Two pure tones presented simultaneously may give rise to combination tones which are frequently called "subjective tones". In contrast to this the two tones presented are sometimes called "primary tones".

PART I

SUBJECT AND METHOD

CHAPTER 2

SURVEY OF PREVIOUS WORK AND MORE DETAILED OUTLINE OF THE PRESENT INVESTIGATIONS

2.1 Introduction

The present investigations deal with two-tone interaction in octave complexes. Reports of previous investigations related to this subject in the literature may be divided into two groups. The first group deals with experiments with two-tone stimuli in which the frequency ratio of the components is slightly different from m:n, where m and n are small integers and $m<n$. These "mistuned consonances" or "mistuned harmonics" can give rise to audible beats.

The second group relates to the experiments with two-tone stimuli in which the frequency ratio of the components is exactly m:n, while the phase relation between the components is adjustable. In this case we will restrict ourselves to the condition $m=1$, since most experiments with phase-locked components and $m\neq1$ described in the literature were concerned with the perception of either third- or higher-order combination tones or the pitch of the stimuli. The present investigations, however, are mainly concerned with the perception of the primary tones themselves (see section 2.5). For $m=1$, the phase variation gives rise to alterations in timbre, loudness, pitch masked threshold of the higher-frequency component, and detectability of the complex in the presence of a masking noise. The alteration in timbre and loudness can concern either the stimulus as a whole or one of its components.

Naturally, the two groups of investigations mentioned above are closely related, since stimuli consisting of two slightly mistuned harmonics can be considered as consisting of two exactly tuned ones with a continuously running phase. Nevertheless, we discuss them separately, because this separation is found throughout the literature.

Each of the two groups comprises certain types of experiments, certain presuppositions and certain explanations of the experimental results which are relevant for the present investigations.

The experiments in the two groups and their results are reviewed and discussed in sections 2.2 and 2.3. The values of L_1 and L_2 at which the various experiments with octave complexes were performed are indicated in Fig. 2.1. The figure will be discussed in section 2.3.4. Section 2.4 contains a survey and discussion of the explanations of the phase effects found in the literature. In section 2.5, the subject of investigation is formulated more precisely on the basis of the above considerations, while finally the main line of our investigations is indicated in section 2.6.

2.2. Previous experiments on beats of mistuned harmonics

The beats evoked by two simultaneously presented pure tones with a frequency ratio slightly different from m:n where m and n are small integers and $m<n$ is a well known phenomenon. The number of beats per second (A) is given by the formula $A=mf_2-nf_1$, published by Ohm in 1839. Plomp (1967) has written an extensive historical review of the discovery of this phenomenon and subsequent studies.

Various authors describe the beats evoked by stimuli with $m=1$ in different ways. Helmholtz (1856) characterized it as a variation in timbre. König (1881) described it as a variation in the loudness of component I. Bosanquet (1879, 1881), Stumpf (1896) and Schouten (1938; see section 2.3) distinguished between a loudness variation of component I and a loudness variation of component II, both at the same rate. These authors carried out their experiments with stimuli in which the frequency of the higher component was 600 Hz or less.

Wegel and Lane (1924) found that the beats heard in stimuli with m=1 were most prominent for a certain value of L_2 relative to L_1. For n=2 and f_1=1200 Hz this was the case when $L_2=L_1-10$ dB. The level conditions for "best beats" were later estimated by Wever (1929), Fletcher (1930), von Békésy (1934), Cotton (1935), Moe (1942), Egan and Klumpp (1951), Opheim and Flottorp (1955), and Lawrence and Yantis (1956a, 1956b) for various values of f_1 and L_1. They all found "best beats" when L_2 was 10 to 20 dB lower than L_1 for n=2. The precise perceptual nature of the beats was not closely investigated by these workers.

Plomp (1967) studied the beats heard in two-tone combinations with m=1 in relation to those in stimuli with $m \neq 1$. Different values of f_1 (125 Hz, 250 Hz, 500 Hz, 1000 Hz and 2000 Hz) and L_1 (between 60 and 100 dB SPL) were considered. The value of L_2 found for "best beats" was plotted as a function of the ratio $\frac{n}{m}$ with L_1 and f_1 as parameters. The resulting graphs show curves which slope downwards with increasing f_2. The experimental points for $m \neq 1$ fit the curves as well as those points for m=1. For n=2, the beats were most pronounced when L_2 was 10 dB lower than L_1. The range of values of m and n for which best beats were found became narrower with increasing f_1. At L_1=100 dB SPL and f_1=125 Hz best beats were found at 25 different ratios m:n, the largest one being 1:12; while at f_1=2000 Hz the maximum ratio at which best beats could be heard was 1:2.

2.3. Previous data on phase effects

2.3.1. *Effect of phase on loudness, timbre and detectability*

Chapin and Firestone (1934) used the octave complex 108/216 Hz with L_1=104 dB SPL as the stimulus. They found that the loudness of this complex was phase dependent for certain values of L_2, with a maximum variation of 12 dB at L_2=87 dB SPL.

Trimmer and Firestone (1937) carried out comparable experiments with stimuli in which the components had a frequency ratio of 1:n with f_1=100 Hz and L_1=104 dB SPL.

Both Chapin and Firestone and Trimmer and Firestone observed variations in timbre (which they called "roughness") as well as loudness variations when the phase was varied. Trimmer and Firestone minimized the loudness and the roughness of the stimuli separately by adjusting L_2 and the phase. With the octave complex 100/200 Hz minimum roughness was reached at L_2=85 dB SPL and minimum loudness at L_2=82 dB SPL. At these values of L_2, the loudness of the stimulus varied about 5 dB as a function of the phase.

Schouten (1938) performed experiments with a frequency combination 200/400 Hz with L_1=105 dB SPL. He found a "very marked" effect of phase on tone quality when L_2 was about 25 dB below L_1.

Craig and Jeffress (1960, 1962) studied the detectability of the difference between the octave complex.

$$p(t)=A_1\sin(2\pi f_1 t) + A_2\sin(2\pi 2f_1 t+\phi)$$

which they called the "basic-phase condition", and the inverse stimulus.

$$p(t)=A_1\sin(2\pi f_1 t) + A_2\sin(2\pi 2f_1 t+\phi+180^{\circ})$$

which they called the "reversed-phase condition".

The value of f_1 in their experiments was 250 Hz, the variables being the phase ϕ, L_1 and L_2. The difference studied could be anything, either the loudness, the timbre or the pitch. On the basis of their results, they divided the phase effects observed into two classes: for $L_1<75$ dB SPL independent of the value of L_2, and for $L_1>75$ dB SPL and $L_2>40$ dB SPL. In the former class the difference between the stimuli was assessed on the basis of timbre and pitch, while in the latter class the assessments were made on the basis of loudness differences.

Similar experiments were performed by Raiford and Schubert (1971) with a method developed by Nixon, Raiford and Schubert (1970). The stimulus here was the frequency combination 250/500 Hz with $L_1=60$ and $L_2=50$ dB SPL. The cosine sum of the frequency components was used as a standard stimulus, and the subjects had to decide whether a second stimulus with a different phase relation differed or not. The scores obtained were found to depend on the phase difference used, with a broad maximum for phase differences round 180°. Further experiments suggested that it does not matter what the phase relation is in the standard stimulus. The latter result was confirmed by Hall and Schroeder (1972), who used the combinations 100/200 Hz, 200/400 Hz and 400/800 Hz as stimuli. They studied the perceptual differences between the stimuli under the various phase conditions by the method of triadic comparison. The level conditions were $L_2=L_1$, $L_2=L_1-6$ dB and $L_2=L_1-12$ dB respectively, with $L_1=50$ dB SPL.

Fricke (1968) studied phase effects from the point of view of signal detection theory (Green, 1958). His frequency combination was 525/1050 Hz, with L_1 and L_2 both 65 dB SPL. After white noise had been added, the detectability of the complex was estimated as a function of the phase for different signal-to-noise ratios near the masked threshold. A four-alternative forced-choice procedure was used. The number of correct responses was found to vary from 25% to 100%, i.e. the whole range of values covered by the psychometric curve.

2.3.2. *Effect of phase on the masked threshold of component II*

Newman, Stevens and Davis (1937) were the first investigators to determine the masked threshold of detectability of component II of an octave complex, with component I as masker. Their stimulus was the combination 370/740 Hz. They measured the masked threshold as a function of L_1. The resulting curve had a slope of about 2 for L_1 between 65 and 80 dB SL. A value of L_1 corresponding to 70 dB SL gave a masked threshold corresponding to 22 dB SL.

The masked threshold as a function of phase was also studied by Clack (1967, 1968), Clack and Bess (1969), Clack, Erdreich and Knighton (1972), Clack (1975), Nelson and Bilger (1974) and De Boer and Bouwmeester (1975). Clack (1967, 1968) and Clack et al (1972) used the frequency combination 1000/2000 Hz as stimulus. The maximum phase effect was 15 dB. Nelson and Bilger (1974) used the combinations 250/500 Hz, 500/1000 Hz, 1000/2000 Hz and 2000/4000 Hz. These authors reported a maximum effect of 12 dB for the higher-frequency combinations and of 33 dB for the lowest-frequency one. De Boer and Bouwmeester (1975) performed experiments with the combination 1000/2000 Hz. Their result shows a maximum effect of nearly 20 dB. All authors performed the experiments at different values of L_1. For the frequency combinations around 1000/2000 Hz and beyond, representative values of L_2 at the masked threshold, averaged over the phase conditions are 30 dB SPL for L_1=70 dB SPL, 50 dB SPL for L_1=80 dB SPL and 70 dB SPL for L_1=90 dB SPL. At lower frequencies, the values of L_2 found are generally lower than the above.

2.3.3. *Effect of phase on pitch*

Very few data on pitch variations as a function of the phase in two-tone stimuli with a frequency ratio of the components of 1:n are available. Stumpf (1910) reported periodic variations of the pitch for frequency ratios slightly different from 1:3 and 1:5. Craig and Jeffress (1962) and Clack (1968) mentioned that their subjects sometimes heard pitch effects when the phase was varied.

Plomp (1967) matched the pitch of the higher-frequency component as a function of the phase for a stimulus with a frequency ratio of 1:3, with f_1=200 Hz. The level conditions were: L_1=100 dB SPL, and L_2 adjusted to the value for which the phase effects were most pronounced ("best beats" condition). He found a pitch shift of maximally 13% upwards. Plomp called this phenomenon the "sweep-tone effect". The curve representing the dependence of pitch on the phase of component II has the shape of a decreasing sawtooth.

A phase-dependent pitch effect in octave complexes was found by Terhardt and Fastl (1971) for the frequency combination 200/400 Hz. In their experiment the value of L_1 was 70 dB SPL, while the value of L_2 was adjusted to 10 dB SL for each phase condition. The maximum pitch shift was 11% upwards.

2.3.4. *Discussion*

The ranges of L_1 and L_2 at which the various experiments discussed above were performed are shown in Fig. 2.1. The initials in the graphs indicate the areas studied by the different authors. The highest frequency combination presented is 500/1000 Hz, in order to avoid a too large variation in frequency combination among the experimental material considered. Two main areas of interest may be distinguished. The first one (shaded in the figure) in which "best beats" could be found, loudness and roughness could be minimized and phase differences were detectable. The second area (bounded by the two broken lines) is that in which the phase-dependent masked threshold can be determined.

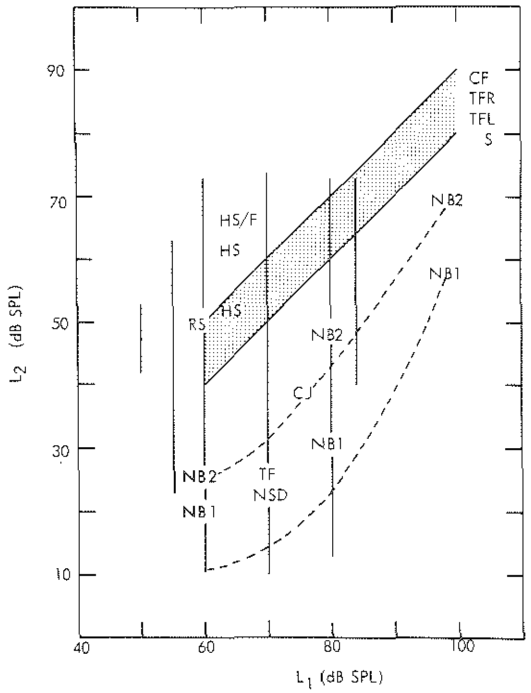


Fig.2.1.

Level conditions under which the different authors mentioned in the present chapter found beats or phase effects in octave complexes. The various level conditions and regions involved are indicated by the initials of the authors concerned. Only results for frequency combinations with f_2<1000 Hz are given. The reader is referred to the various sections of this chapter for a description of the experiments in question. The shaded area indicates the conditions under which "best beats" were found. The vertical lines indicate the conditions under which Craig and Jeffress (1962) detected differences between a given stimulus and the reverse one. The significance of the initials is as follows:

Initial	Authors	Topic Studied
CF	Chapin and Firestone (1934)	Minimum loudness
TFR	Trimmer and Firestone (1937)	Minimum roughness
TFL	Trimmer and Firestone (1937)	Minimum loudness
NSD	Newman, Stevens and Davis (1937)	Masked threshold
S	Schouten (1938)	Minimum roughness and beats
F	Fricke (1967)	Differences in detectability
RS	Raiford and Schubert (1971)	Differences in timbre
HS	Hall and Schroeder (1972)	Differences in "whatever might be"
NB1/NB2	Nelson and Bilger (1974)	Masked threshold averaged over the phase conditions. NB1 corresponds to the combination 250/500 Hz and NB2 to the one 500/1000 Hz
TF	Terhardt and Fastl (1971)	Pitch

The initials CJ within a square refer to the division in classes of perceptual phenomena by Craig and Jeffress (1962; see section 2.3.1). The two broken curves represent very roughly the upper and lower boundaries of the masked threshold of component II depending on phase.

Unfortunately, the phase effects do not give rise to a uniform sensation in each of these two areas. For example, Trimmer and Firestone (1937) described the phase effects in the "best beats" area as variations in loudness and timbre, while Craig and Jeffress (1962) heard only variations in loudness in the same region. In the other region, Terhardt and Fastl (1971) found a phase-dependent pitch shift for L_2=10 dB SPL (L_1=70 dB SPL), while Nelson and Bilger (1974) found no such shift and Craig and Jeffress (1962) reported that the detection was based on timbre and pitch differences. We may thus conclude that the sensations observed in the regions which have been investigated so far are very complex. On the other hand, very few studies have been performed in the area between the two mentioned above. This region, therefore, would seem to be a promising one for new investigations.

2.4. Explanations proposed for the phase effects

2.4.1. *Introduction*

The two main explanations proposed for the origin of the phase effects (including beats) for the frequency ratio 1:n are as follows:

(1) Component I of the complex, if presented at a sufficiently high level, gives rise to subjective ("aural") harmonics generated by distortion in the mechanical/hydraulic part of the auditory system (the middle-ear system or the cochlea). The phase effects are supposed to be due to interference between the aural harmonic and the highest frequency component of the stimulus.

(2) The phase effects are related to variations in the waveform of the superimposed sinusoidal components.

A third explanation found in literature relates the phase effects to interference between a combination tone

("f_2-f_1" for n=2) and component I of the stimulus.

This explanation will not be considered here, since we are mainly interested in the perception of the higher-frequency component (see section 2.5).

Supporters of both hypotheses can be found in the literature. Plomp (1967) has given an extensive review of the various points of view expressed.

2.4.2. *Aural harmonics*

The idea that aural harmonics might cause phase effects was first suggested by Müller (1871) and Hermann (1896) (see Plomp, 1967). Many investigators, already mentioned in section 2.2, used the method of "best beats" for frequency ratios of the components slightly different from 1:n to estimate the levels of the aural harmonics - which were thought to correspond to the levels of the higher-frequency component for which the beats were most pronounced. The values of L_2 found in this way were very high, *viz.* 10 to 20 dB below L_1. All investigators regarded the presence of the beats and the possibility of adjusting the experimental conditions so as to give "best beats" as demonstrations of the existence of aural harmonics.

Plomp (1967) performed experiments in order to determine whether the beats under the "best-beats" conditions are actually caused by aural harmonics. His reasoning was based on the monotonic shape of the curve representing the value of L_2 for best beats as a function of $\frac{n}{m}$ for a given value of L_1.

As we have already mentioned in section 2.2, the values for $m\neq1$ fitted this curve as well as those for $m\neq1$. Plomp added high-pass noise to a stimulus with $m\neq1$ in order to mask the supposed aural harmonics. The beats remained audible so they could not be caused by interaction of the n^{th} harmonic of component m and the m^{th} harmonic of component n.

Plomp extrapolated this result to the case m=1. He concluded that, at lower frequencies, the beats were not caused solely by aural harmonics. The extrapolation procedure failed at higher frequencies because e.g. at f_1=2000 Hz beats could only be heard for n=2 and m=1. In these cases, a possible rôle of aural harmonics could not be excluded. Furthermore, Plomp argued that the sweep-tone effect in the pitch is difficult to explain in terms of nonlinear distortion.

Plomp (1967) also checked whether aural harmonics could be heard in a pure-tone stimulus, and found that this was sometimes - but by no means always - the case. Moreover, there was no relation between this ability and the ability to hear beats.

A second method often used to detect subjective harmonics is the steady-tone phase-effect method (see section 2.3.1). The value of L_2 at which the loudness of the complex is minimum under specific phase conditions has been assumed to correspond to the level of the aural harmonic. This interpretation was complicated by the fact that not only loudness but also timbre ("roughness") varies with the phase. For this reason Trimmer and Firestone (1937) concluded that component II of the complex cannot be considered as an independent "probe" tone. They considered it "pointless to try to use the phase effect as a probe method for studying the phases and amplitudes of the subjective harmonics of the fundamental". Comparable reserve can be found in the paper by Schouten (1938).

The aural-harmonics explanation was also supported by Clack and co-workers (see section 2.3.2), Schubert (1969) and Erdreich and Clack (1972), on the basis of the results of their experiments on the phase dependence of the masked threshold of component II for the frequency combination 1000/2000 Hz. The authors examined whether vector summation of the supposed aural second harmonic and component II agreed with their results. The calculated levels of the assumed aural harmonic, plotted as a function of L_1, gave a straight line with a slope of 2. A lower-frequency component with L_1=70 dB SPL gave rise to a supposed aural second harmonic with a level corresponding to 20 dB SPL. Nelson and Bilger (1974) found that this slope was frequency-dependent; for example, at f_1=250 Hz a slope of 1 was found.

De Boer and Bouwmeester (1975) applied the vector-summation model to results obtained with the same frequency combination as that used by Clack and co-workers. The levels calculated were similar to those of Clack and co-workers. De Boer and Bouwmeester also calculated the phases of the assumed aural harmonics which, however, varied widely from observer to observer. No level dependence was found.

2.4.3. *The waveform hypothesis and other explanations*

De Morgan (1864) was probably the first to prove that the rate at which the waveform obtained by superimposing two sine waves with a frequency ratio slightly different from m:n varies is in agreement with the formula of Ohm (1839; see Plomp, 1967). The waveform hypothesis was further adopted by König (1876), Lottermoser (1937), Meyer (1949, 1954, 1957), Chocholle and Legouix (1957a, 1957b) and Plomp (1967).

These authors' preference for the waveform hypothesis was generally based on rejection of the aural-harmonics explanation on the basis of two arguments. Firstly, it was claimed that if the aural harmonics were as strong as suggested by the "best-beats" method, they should be audible in a pure tone. Secondly, experiments with three harmonically related components showed that when the components were slightly mistuned, beats could be heard even at very low levels (Thomson, 1878; Ter Kuile, 1902), which makes it improbable that these are caused by nonlinear distortion.

The experimental evidence against the aural-harmonics explanation presented by Plomp (1967) has already been described in section 2.4.1. Plomp suggested that the beating sensation might be caused by slow variations in the time pattern of the nerve impulses evoked by the stimulus. The sweep-tone effect might be understood from the point of view of the "time theory", in which the reciprocals of the time intervals between successive prominent peaks in the waveform are thought to represent the pitch (Schouten et al., 1962; Ritsma, 1967). Plomp considered the time pattern of the waveform of a stimulus with $f_1:f_2 = 1:n$ as consisting mainly of two pulse trains, with a time delay between the two that varies with the phase. The reciprocal of this time delay varies like a sawtooth as a function of phase.

Craig and Jeffress (1962) took a reserved stand upon both the aural harmonics and the waveform hypothesis. They proposed a third possible explanation for the phase effects, according to which these are thought to be caused by a shift in the locus of vibration on the basilar membrane. This hypothesis was based on ideas of von Békésy (1957) originating from his experiments with models of the cochlea.

2.4.4. *Discussion*

Our survey of the literature has shown that neither the aural-harmonics hypothesis nor the waveform hypothesis has achieved general acceptance as an explanation of the phase effects. We feel that one of the main reasons why a generally acceptable model has not been developed so far is the lack of attention to the perceptual nature of the beats. We need to be much better aware of which perceptual attributes of the tonal complex actually vary when the phase is changed.

A similar objection was made e.g. by Trimmer and Firestone (1937; see section 2.4.2), who concluded that component II of the complex could not be used as an independent probe tone, because of the differences in timbre involved.

The study of Plomp (1967; see section 2.4.2) has provided quite strong evidence against the validity of the aural-harmonics hypothesis. The fact that beats remained audible even though the presumed aural harmonics had been masked, may be regarded as a strong indication that beats under "best beats" conditions at low frequencies are not due to aural harmonics.

Another crucial point in this connection is whether the vector summation model (see section 2.4.2) is applicable, i.e. whether the beats are really due to vector summation of the aural harmonics and component II of the complex. The results of Clack and co-workers and of de Boer and Bouwmeester (section 2.3.2) seem to suggest that this model is applicable, in view of the good fit of their calculated results with the experimental points, and in particular in view of the fact that the curve of the level of the presumed aural harmonic as a function of L_1 has a slope of 2, as predicted. On the other hand, Nelson and Bilger (1974) considered the fact that the magnitude of the effect of phase on the masked threshold, and the slope of the masked threshold-L_1 curve, were frequency-dependent was an argument against the assumption of aural harmonics produced by distortion in the mechanical/hydraulic part of the ear.

As we have already mentioned, there is little direct experimental evidence for the waveform hypothesis. Most of the authors who have expressed themselves in favour of this hypothesis have done so because the aural-harmonics hypothesis had, in their opinion, been discredited by evidence such as that given above.

It will thus be clear that there is an urgent need for experimental evidence that will help us to make a definite choice between the various alternative explanations. The main points which require attention are the perceptual nature of the phase effects themselves, whether component II can be considered as an independent probe tone and whether the vector-summation model is applicable.

2.5. The scope of the present investigations

We concluded in section 2.3.4 that very few experimental data are available in the intermediate region below that for best beats and above that for the masked threshold of component II (see fig. 2.1). Preliminary listening revealed that the phase effects in this region are relatively simple from a perceptual view, since only the loudness and pitch of the higher-frequency component were involved. The sensations in this region were, therefore, chosen as our primary field of investigation.

We may note that at higher values of L_1, this intermediate region becomes narrower because the masked-threshold curves become steeper than the "best beats" curves. This effect is still more distinct for higher values of f_1, as shown by the results of Clack et al. (1972). However, the masked threshold will only play an incidental part in our investigations.

There is reason to expect that study of the pitch and loudness effects near the masked threshold may help us to determine whether component II may be considered as an independent probe tone, and whether the vector-summation model is applicable under these conditions. Finally, we hope to find direct evidence for or against the waveform hypothesis.

Summarizing, we hope to provide an answer to the follow-

ing four questions with our investigations:
(1) What are the sensations of the stimulus components in the field of investigation chosen?
(2) May the higher-frequency component of the complex be considered as an independent probe tone?
(3) Is the vector-summation model applicable, i.e. do the calculated levels of the aural harmonics really correspond to components present in the ear?
(4) Is there experimental evidence for the waveform explanation?

2.6. Outline of the present investigations

(1) The first topic of investigation is formed by the phase effects in the perception of octave complexes for values of L_1 beyond 60 to 70 dB SPL and L_2 at least 20 dB below L_1. Special attention will be paid to possible pitch effects in this area. The effects observed with octave complexes are then set in a framework of interaction phenomena in two-tone stimuli in general, with or without a phase relation between the components.
(2) Extensive pulsation-threshold experiments will be described. As we have already mentioned in Chapter 1 and will discuss in greater detail in Chapter 3 the pulsation threshold reflects some kind of internal representation of the stimuli. This method may, therefore, help us to study how the stimuli are processed. We are interested in the test-tone frequency region involved in the interaction, the dynamic behaviour in that region, nonlinearities and possible phase relations. In addition, the questions formulated in section 2.5 will be discussed in Chapters 8 and 9.
(3) The contents of Chapter 10 forms a preparation to a model to be presented in Chapter 11. We shall consider the dynamic behaviour of the pulsation-threshold results in relation to the results of forward-masking experiments and

discuss the implications of this comparison for the interpretation of the phase effects. The model to be presented in Chapter 11 may be regarded as a first step in seeking for an alternative to the aural-harmonics and waveform hypothesis.

CHAPTER 3

METHODS, PROCEDURES AND APPARATUS

3.1. Methods and procedures

3.1.1. *Pitch experiments*

In the pitch experiments component II was presented alternately with a pure-tone stimulus in the sequence indicated in Fig. 3.1a. The observer adjusted the pure tone in frequency until component II and the pure-tone stimulus had the same pitch. This adjustment was performed by turning a knob, starting from an arbitrary initial value of the frequency. Before each adjustment the level of the pure-tone stimulus was adjusted so that the loudness of this stimulus was roughly equal to that of component II.

The duration of presentation of component II varied from 400 Hz for the lowest frequencies investigated to 200 ms for the highest ones. Component II was presented with a shorter duration than component I, to facilitate the recognition. The pure-tone stimulus which was adjusted in frequency is sometimes called the "matching tone". The matching tone had the same duration as component II. The pitch is expressed in relative terms by the ratio $\frac{f_m-f_2}{f_2}$, where f_m is the adjusted frequency of the matching tone.

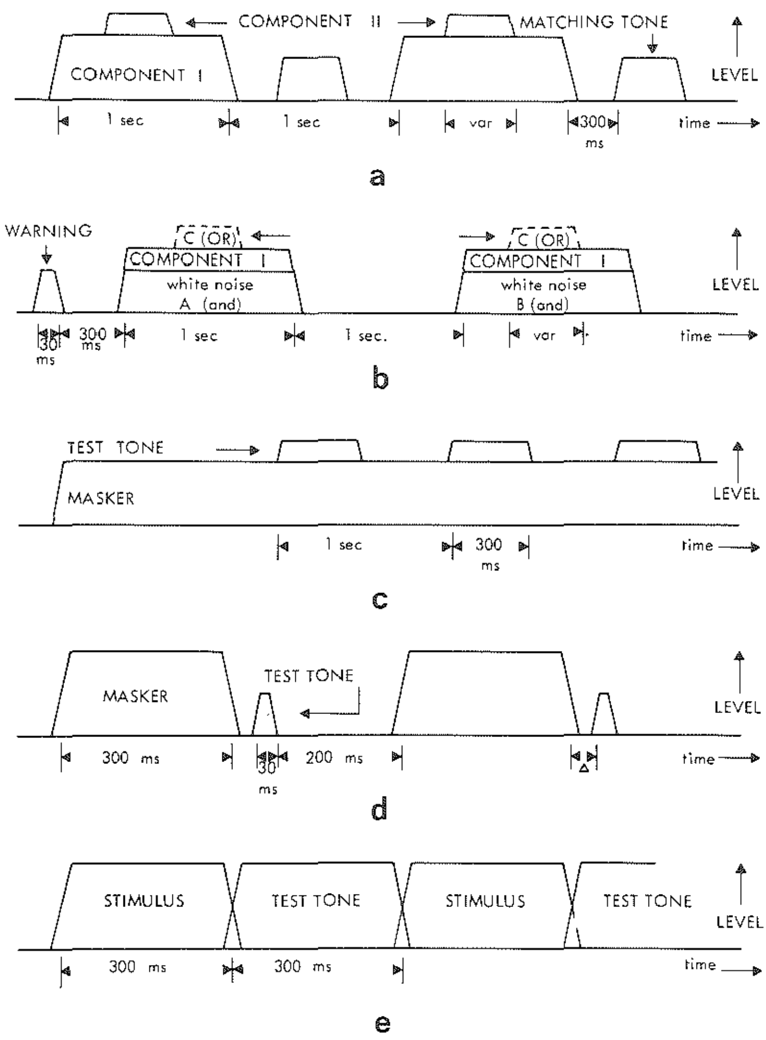


Fig. 3.1.

Stimulus configuration in the
a) pitch and loudness matchings;
b) masking-level experiments;
c) simultaneous-masking experiments;
d) forward-masking experiments;
e) pulsation-threshold experiments.

3.1.2. *Loudness experiments*

The same stimulus configuration as for the pitch experiment was used to match the loudness of component II (Fig. 3.1.a). The observer adjusted the pure-tone stimulus in level until it had the same loudness as component II. This adjustment was again performed by turning a knob, starting from an arbitrary initial level. Before each adjustment the pitch of the matching tone was made roughly equal to that of component II. The loudness of component II is expressed by the level of the matching tone.

3.1.3. *Masking-level measurements*

In measurements of the masking level, the lowest level of white noise needed to mask a given component of a complex sound is taken as a measure of the subjective strength of the latter. A linear relation exists between the level of a pure-tone stimulus and the above mentioned noise level, apart from some departure from linearity near the absolute threshold (Hawkins and Stevens, 1950). In the present experiments, the method was used to measure the strength of component II of octave complexes. The subjects determined the white-noise masking level in a two-alternative forced-choice (2-AFC) procedure. The configuration of the stimuli is shown in Fig. 3.1.b. Each trial was preceded by a pure-tone pulse that acted as a warning signal, the pitch and the loudness of which were roughly equal to those of component II. Two points of the psychometric curve were determined by adjusting the noise level, and the noise level corresponding to 75% correct responses was obtained by interpolation. This value was called the "white-noise masking level". This procedure was not followed under those phase conditions where component II was near the masked threshold. In these cases the detectability of the signal was measured in the absence of the noise, and the psychometric curve for a pure tone of frequency f_2 was used to transform the percentage detectability into decibels with respect to the absolute threshold, expressed in terms of the white-noise masking level. This threshold was defined with the aid of the Hawkins and Stevens function just mentioned. The linear part of this curve was extrapolated until it cut the level corresponding to the absolute threshold of the tone. The white-noise masking level corresponding to this point was called the absolute threshold in terms of the masking level.

3.1.4. *Simultaneous-masking experiments*

In the simultaneous-masking procedure the masker and the test tone are presented simultaneously. The sequence of presentation we used is indicated in Fig. 3.1.c. The simultaneous-masked threshold of the test tone was defined as the highest level of this tone for which the observer could not decide whether the test tone was present or not. The experiments were performed by an adjustment procedure. In our experiments, component I was used as masker and component II as test tone. This type of experiment is called "pure-tone masking".

3.1.5. *Forward-masking experiments*

The forward-masking method is a non-simultaneous masking procedure in which the test tone is affected by the masker shortly after the termination of the latter. One makes use of the residual adaptation left after the masker is switched off. The temporal stimulus configuration is given in Fig. 3.1.d. The forward-masked threshold of the test tone was defined as the highest level for which it was not possible to decide whether the test tone was present or not. Most of the experiments were performed by a 2-AFC procedure, a few by an adjustment procedure.

3.1.6. *Pulsation-threshold experiments*

In the pulsation-threshold method the stimulus and the test tone are presented alternately. The sequence of presentation is indicated in Fig. 3.1.e. The pulsation threshold of a test tone was obtained by adjusting its level to the highest value at which the test tone appeared to be continuously present. Because of the alternating presentation of stimulus and test tone, the pulsation-threshold method is often considered as a non-simultaneous masking procedure (see e.g. Houtgast, 1973).

3.2. Discussion of the masking methods

Masking experiments are used to study the internal representation of sounds (see Chapter 1). As we mentioned above, there are two types of masking methods, simultaneous masking and non-simultaneous masking. The simultaneous method has been used most frequently in the literature (Chapter 2). The pulsation-threshold method, introduced recently by Houtgast (1972)*, is not really a masking method at all, as the test tone remains audible below the "threshold". Houtgast sometimes called the pulsation threshold of the test tone the masking effectiveness of the stimulus in the frequency region of the test tone. In view of the analogy with simultaneous masking we shall continue to speak of "masker" and "test tone" even when the pulsation-threshold method is being applied.

The second non-simultaneous procedure, forward masking, makes use of the residual adaptation left after the masker is switched off. The pulsation-threshold methods and forward masking lead to very comparable results. We shall return to this point in Chapter 10.

*This method is based on the "continuity effect" discovered by Thurlow (1957).

Houtgast (1973) made an extensive comparison of the simultaneous and non-simultaneous procedures. The results obtained with the non-simultaneous procedures revealed a specifically nonlinear phenomenon. Under these conditions, the masking evoked by a pure tone could be reduced by adding other sounds to the stimulus. The phenomenon has been called "two-tone suppression" or "two-tone inhibition". The same phenomenon has been found in the results of single-cell recordings in the auditory nerve (Kiang et al.,1965 and Sachs and Kiang, 1968). The phenomenon of suppression has never been found with the simultaneous-masking procedure. This has been interpreted as due to the fact that the test tone passes the same suppressing nonlinear mechanism as the stimulus, implying that the signal-to-noise ratio determining the threshold of detectability is measured at a stage prior to the nonlinearity. We, therefore, prefer non-simultaneous masking procedures, in particular the pulsation-threshold method because it is more sensitive than the forward-masking procedure. Simultaneous-masking experiments have only been performed incidentally, to permit a comparison with data given in the literature.

3.3. Apparatus

The block diagram of the apparatus we used in our investigations is given in Fig. 3.2. Each phase-lock generator consists of a Hewlett and Packard 3300 A function generator in combination with a 3302 A trigger phase-lock plug-in unit. The first generator was triggered directly by an oscillator and the second one after division of the frequency by two. The output signals of the two phase-lock generators for the stimulus and another oscillator for the test tone were filtered (to suppress harmonic distortion) and gated. The gates were part of a preset-counter/gate circuit which enables the experimenter to adjust the required stim-

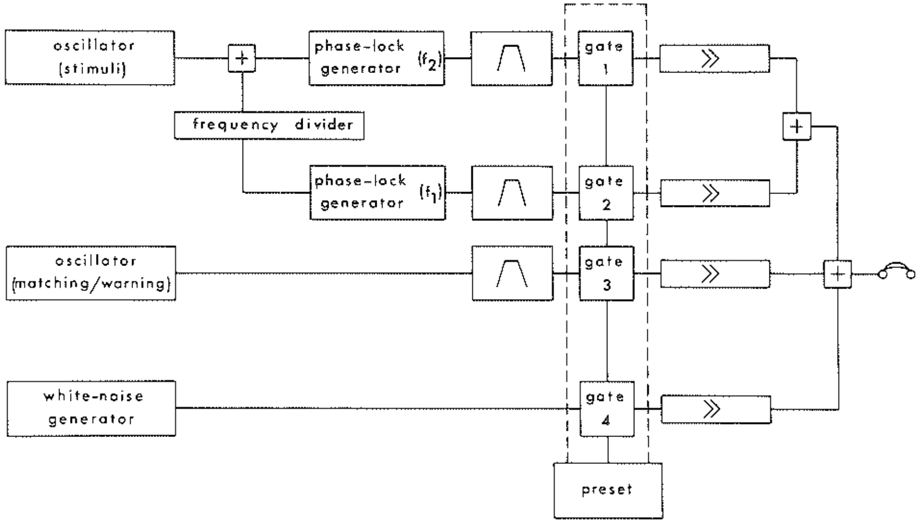


Fig. 3.2.

Block diagram of the apparatus

ulus configuration, including the configuration for the forced-choice procedure.

The signals presented were given a Gaussian envelope, having a 10-90% rise/decay time of 20 ms. The duration of the signals was defined as the time between the initial and final half-height points. The second harmonic distortion of the tones presented was below -55 dB. A Grason Stadler TDH 39 circumaural earphone was used. The observers were seated comfortably in an acoustically insulated booth.

The stimuli were presented monaurally. A total of ten observers participated in the experiments. Most experiments were performed by at least two observers.

3.4. Presentation of the results

The results are presented throughout this study as a function of the phase of component I. A full cycle of the different waveforms of the complex then corresponds to a phase range from 0° to 180°.

The phase relation between the acoustic signal and the electric signal on the earphone was estimated as a function of frequency with a Bruel and Kjaer 4153 artificial ear, without the cushion. With the aid of this phase characteristic the stimulus can be represented by the formula

$$f(t)=A_1\sin(2\pi f_1 t) + A_2\sin(2\pi 2f_1 t+\phi_{2a}) \qquad (1)$$

where ϕ_{2a} corresponds to the phase of component II in the acoustic signal as determined with the artificial ear. Formula (1) can be re-written in the form

$$f(t)=A_1\sin(2\pi f_1 t+\phi_{1a}) + A_2\sin(2\pi 2f_1 t)$$

where $\phi_{1a}=-\frac{1}{2}\phi_{2a}$, represents the phase of component I in the acoustic signal. For the sake of surveyable lay out of the diagrams, the results are in fact plotted as a function of the phase ϕ_1 of component I where

$$\phi_1=\phi_{1a}+15^\circ$$

To facilitate the conversion of ϕ_1 into ϕ_{2a}, a conversion scale is given in Fig. 3.3.

We shall discuss in Chapter 7 how the phase relation between the acoustic and electric signals is influenced by the earphone cushion, and how the phase shift in the ear canal and the middle ear can be taken into account.

Fig. 3.3.

Conversion scale between ϕ_1, the phase plotted in the presentation of the experimental results, and the phase ϕ_{2a} of the higher-frequency component in the acoustic signal at the entrance of the ear canal. See also section 3.4.

ϕ_1	ϕ_{2a}
0°	30°
30°	330°
60°	270°
90°	210°
120°	150°
150°	90°
180°	30°

PART II

QUANTIFICATION OF THE SENSATIONS INVOLVED IN THE PERCEPTION OF PHASE EFFECTS IN TWO-TONE OCTAVE COMPLEXES

CHAPTER 4

PERCEPTION OF THE HIGHER-FREQUENCY COMPONENT IN TWO-TONE OCTAVE COMPLEXES

4.1. Phase dependence of pitch, masking level and loudness

The pitch, masking level and loudness of component II were measured as a function of phase for the frequency combinations 200/400 Hz, 400/800 Hz, 760/1520 Hz and 1600/3200 Hz. Three observers participated in the experiments. Some representative results are shown in Fig. 4.1. and 4.2. Each panel in these figures gives the results for a specific (L_1,L_2) combination. The phase dependence of the masking level and the pitch is given in Fig. 4.1. for the four (f_1,f_2) combinations. The phase depencence of the loudness and the pitch is given in the Figs. 4.2.a and 4.2.b for the (f_1,f_2) combination 200/400 Hz and in the Figs. 4.2.c and d for the combination 400/800 Hz. For the 1600/3200 Hz frequency combination, none of the observers could detect a pitch shift.

The shape of the masking-level and loudness curves was found to be remarkably constant for a specific (f_1,f_2) combination. The parts of the masking-level curves at the low-phase side of the minimum are generally somewhat flatter than those at the high-phase side. This is very pronounced

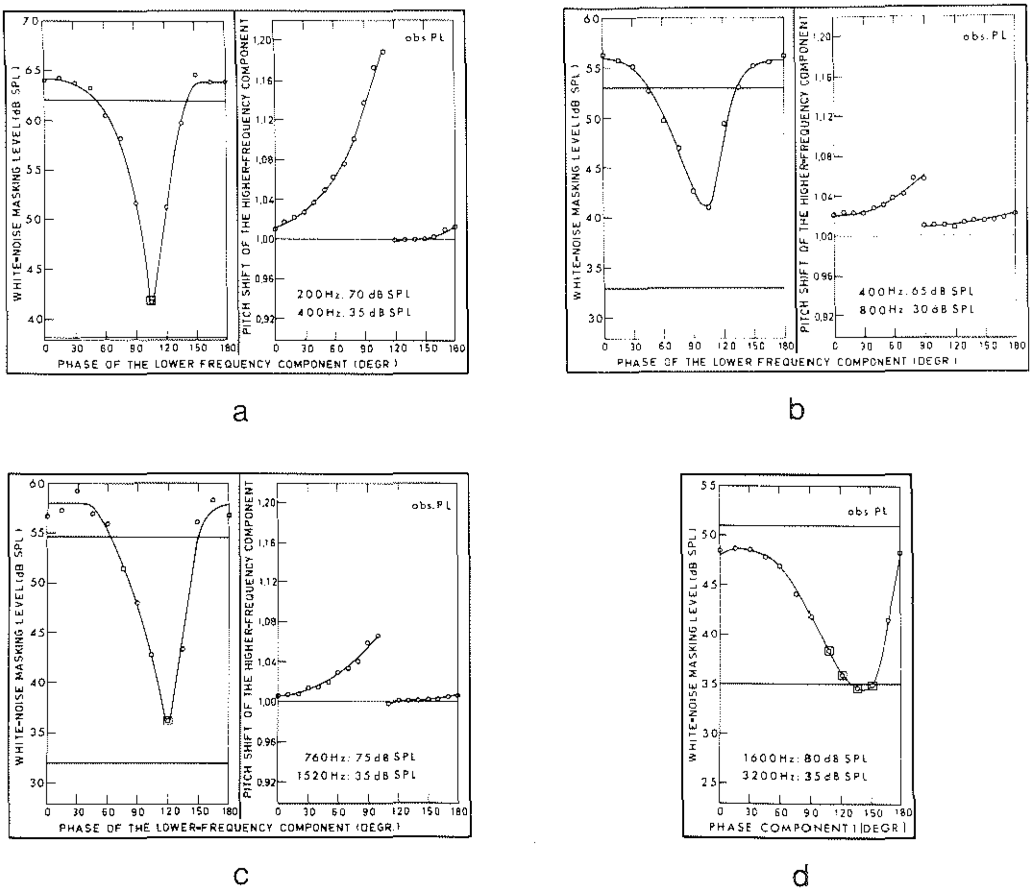


Fig. 4.1.

White noise masking level and pitch of component II as a function of the phase of component I (ϕ_1) for different frequency and level combinations. The upper horizontal lines in the graphs represent the masking level of component II in the absence of component I. The lower horizontal lines represent the absolute threshold of component II. For the definition of this threshold in terms of the white-noise masking level we refer to section 3.1.3. The experimental points of the masking level near the absolute threshold of component II (indicated by ⊡) were measured as percentage detectabilities in the absence of the noise and transformed into dB with respect to the absolute threshold expressed in terms of the white-noise masking level (see section 3.1.3). The pitch of component II is plotted in the right hand half of each graph as the frequency ratio of the matched pure tone and component II. The horizontal lines in the pitch graphs correspond to the pitch which would have been obtained in the absence of component I.

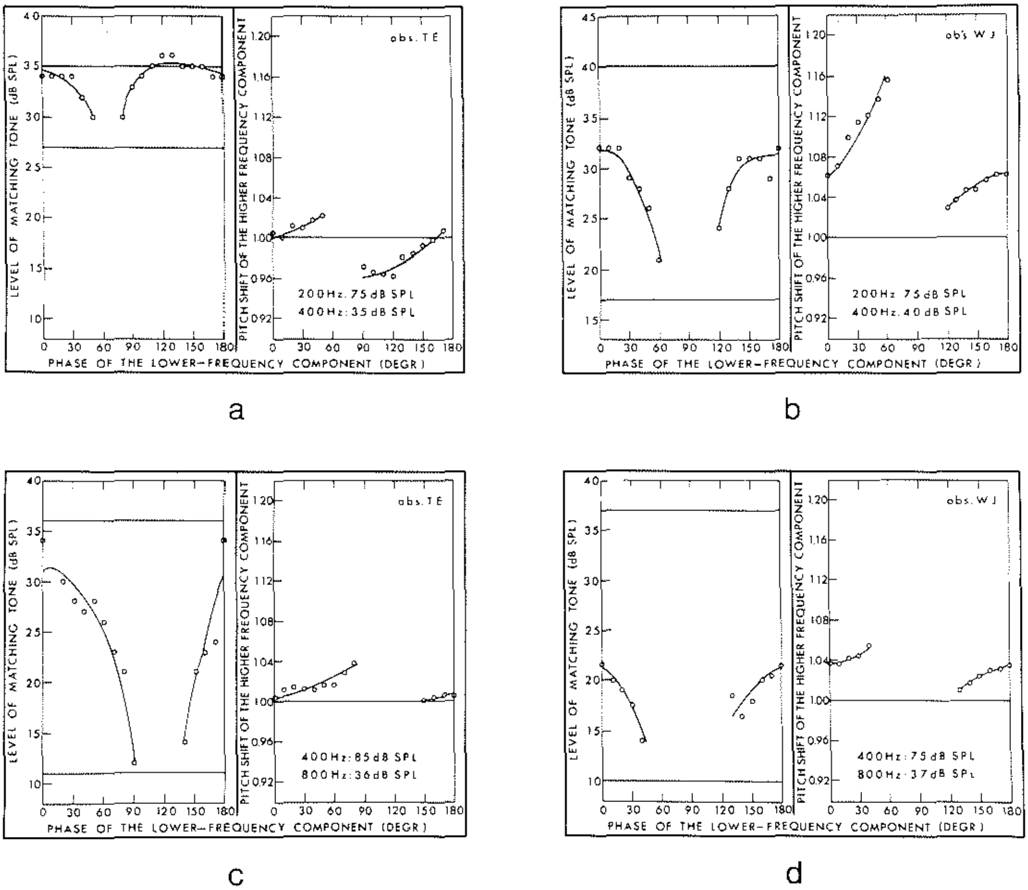


Fig. 4.2.

Loudness and pitch of component II as a function of the phase of component I (ϕ_1). The significance of the horizontal lines in the graphs is the same as in Fig. 4.1.

for the 1600/3200 Hz combination. The gap in a number of curves for some phase region means that component II was not detectable at all. The maximum phase effect found in the masking-level and loudness results is 20 dB.

Under the phase condition corresponding to maximum audibility of component II the masking level is generally a few decibels higher than that for component II alone. A comparable effect for the loudness was not found. In fact, the loudness was ofter lower than that of component II alone under these conditions.

The pitch-phase curve generally resembles a sawtooth function. The pitch is roughly equal to that of component II alone at the beginning of the sweep and higher at all other phases. The maximum pitch shift found is 20%. Under some conditions one observer (TE) obtained curves that started at a lower pitch than that of component II alone (Fig. 4.2.a). For a specific (f_1,f_2) and (L_1,L_2) combination (400/800 Hz, 82/33 dB SPL), observers TE and PL found ambiguity in the pitch of component II; The results for observer TE are shown in Fig. 4.3. All observers reported that component II of the complex sounded like a pure tone, but with a pitch depending on the phase.

The scatter in the masking-level and loudness data depends on whether the masking level and loudness are close to that of component II alone or whether they correspond to the absolute threshold of component II. This scatter (defined as twice the standard deviation for three adjustments increases from 2 dB at higher levels and loudnesses to 5 dB near the threshold. The accuracy of the pitch matchings varies in a similar way between 0.3% and 2%.

The results show an additional day-to-day variation in the depth of the minimum in the curves. This subject will be discussed in section 4.3.2.

The phase of the minimum in the masking-level and loudness curves coincides with that of the jump in the corresponding pitch curves. The phase at which the minimum occurs is almost constant for each (f_1, f_2) combination. Closer investigation of this dependence for the 400/800 Hz frequency combination revealed an increase in the phase of this

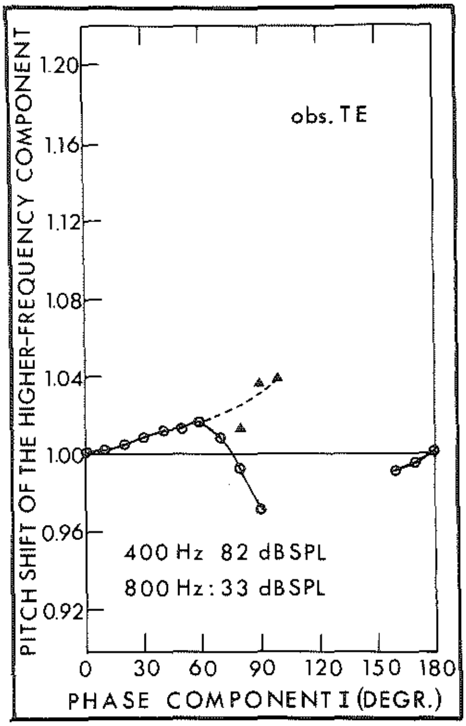


Fig. 4.3.

Pitch of component II as a function of ϕ_1. The different symbols indicate the different pitches that could be found under specific phase conditions.

minimum with increasing value of L_1 (about 10^o/10 dB). No dependence on L_2 was found. The scatter in the phase adjustments is 20^o. The dependence of the phase for minimum loudness on the absolute frequencies of the components will be described in Chapter 7. The highest frequency combination for which we found a loudness minimum is 3500/7000 Hz.

Both the pitch and the loudness effects could also be perceived when the stimuli were presented binaurally. However, with dichotic presentation of a slightly mistuned octave complex (one tone to each ear), the observers only had a very vague sensation of something moving inside the head.

4.2. Phase effects as functions of L_1

The pitch, the masking level and the loudness of component II were then measured as a function of L_1. We only investigated the two extreme phase conditions i.e. those corresponding to minimum audibility of component II (ϕ_{min}) and to maximum audibility (ϕ_{max}). The measurements were performed with L_2 as the parameter, the maximum value of L_2 being 45 dB SPL. The (f_1,f_2) combinations were the same as in the preceding section. The value of L_1 was increased in steps of 1 or 2 dB over a range of up to 30 dB, until component II was fully masked. Typical results are shown in Fig. 4.4 for pitch, in Fig. 4.5.a for masking level and in Fig. 4.5.b for loudness.

When $\phi_1=\phi_{min}$ the pitch of component II increases very fast at higher values of L_1. When $\phi_1=\phi_{max}$ the pitch does not change with L_1 but is generally somewhat higher than in the absence of component I. The maximum pitch shift evoked by component I and the shape of the curves do not depend on L_2. When $\phi_1=\phi_{max}$, there is a change in timbre at the same time as the decrease in loudness. No change in timbre was found for $\phi_1=\phi_{min}$. The scatter in the adjustments varies from 0.3% in cases with little pitch shift up to 1-2% for the maximum pitch shift.

The masking level and the loudness fall off sharply just before component II is fully masked by component I. The increase in the masking level above that in the absence of component I has already been described in section 4.1.

Each pair of curves shown in Fig. 4.4 and 4.5 was sometimes shifted in a horizontal direction by a few decibels when determined in different sessions. However, the distance between the two curves remained constant. This variation is not due to differences in the position of the earphone.

Naturally, component II could not be detected and no difference could be heard between the complex and component I alone if the curves for $\phi_1=\phi_{min}$ (120°) in Fig. 4.5.a and

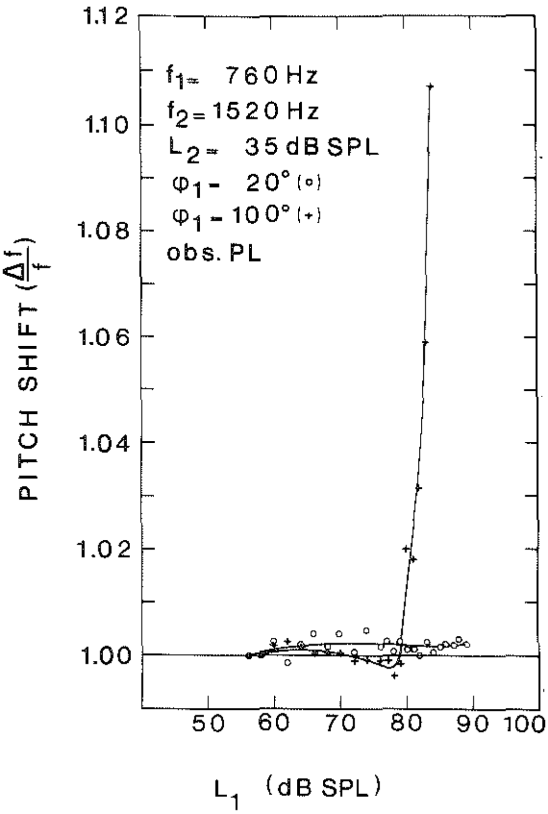


Fig. 4.4.

Pitch of component II of the octave complex 760/1520 Hz as a function of L_1 under the two extreme phase conditions, corresponding to maximum and minimum audibility of component II.

4.5.b crossed the absolute threshold level. However, a difference could be heard again if L_1 was increased slightly. All observers had the impression that component I was involved in the perceptual alteration due to the addition of component II to component I for this (L_1,L_2) combination. The effect was most marked for the (f_1,f_2) combination 200/400 Hz*.

*This effect indicates that different detection criteria can be used in the determination of the masked threshold. Smoorenburg (1972) differentiated between the threshold of identification and the threshold of detection. The threshold of identification corresponds to the disappearance of the characteristic pitch of the test tone, while the threshold of detection is found when no difference can be heard between the masker with and without the test tone; interaction phenomena such as combination tones then determine the masked threshold. In our experiments we determined the identification threshold.

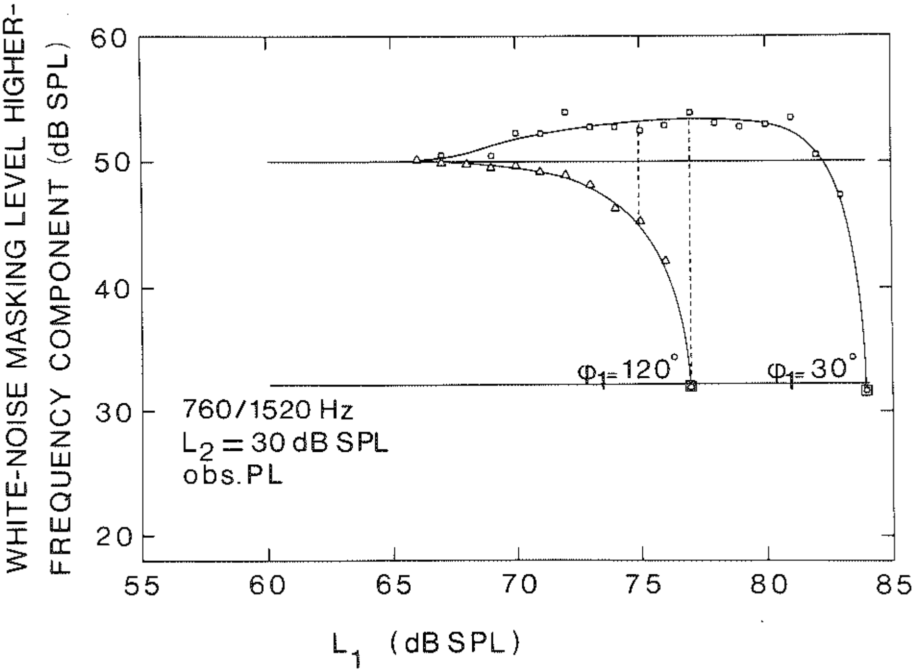


Fig. 4.5.a.

Masking level of component II as a function of L_1 for $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$. The upper horizontal line indicates the masking level of component II alone, and the lower one the absolute threshold of component II. The experimental points within squares near the absolute threshold were measured as percentage detectabilities in the absence of noise (see section 3.1.3). The broken vertical lines are explained in section 4.3.2.

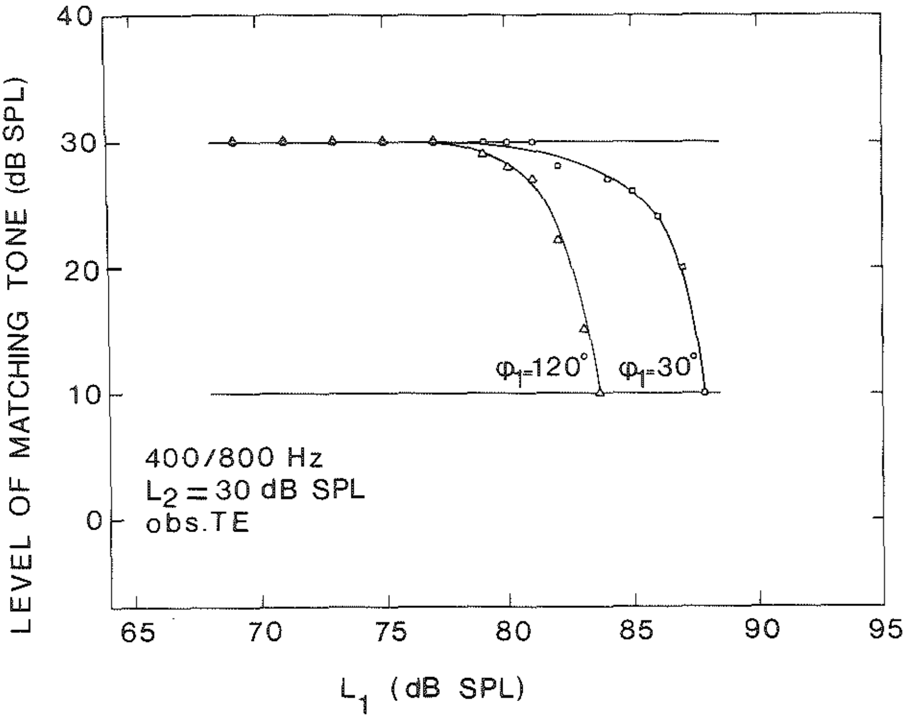


Fig. 4.5.b.

Loudness of component II as a function of L_1 for $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$. The upper horizontal line indicates the loudness of component II alone and the lower one the absolute threshold of component II.

4.3. Discussion

4.3.1. *Pitch*

The results presented show a distinct change in the perceived pitch of component II with changing phase.

The maximum pitch effect found (20%) agrees well with the results of Terhardt and Fastl (1971) who found a maximum effect of 11% for the (f_1,f_2) combination 200/400 Hz. However, the shapes of the curves are not comparable, owing to the different methods of measurement used.

The sawtooth-like shape of our curves agrees very nicely with the results of Plomp (1967) for a 1:3 frequency ratio (200/600 Hz). Plomp plotted his results as a function of the phase of the higher-frequency component. If we mirror his figure, i.e. plot his results as a function of the phase of the lower-frequency component, we get a sawtoothlike curve with a positive slope just as in our results. The agreement seems to indicate a similar interaction mechanism for the two frequency ratios.

One might expect the change in pitch as a function of phase to be due to the change in loudness. There are two reasons why we do not believe this to be the case. First, the pitch shifts we found are much larger than can be expected on the basis of a pitch-loudness dependence (maximum 2%/20 dB; Verschuure, 1975). Secondly, the curve representing the dependence of pitch on phase is not symmetrical with respect to the phase giving maximum pitch shift. Equal loudnesses would give rise to equal pitches.

4.3.2. *Masking level and loudness*

It may be concluded from Fig. 4.2. that the loudness of component II when $\phi_1=\phi_{max}$ tends to be matched lower than when component II is matched alone. Zwicker (1963), Scharf (1964) and Houtgast (1974a) found that the loudness of a tone can be reduced in the presence of a noise. Houtgast (1974a) correlated this "loudness reduction" with the phenomenon of "two-tone suppression".*

*To avoid confusion between the terms "suppression" and "reduction", we define them as follows. "Suppression" is the phenomenon that for a pure-tone stimulus the pulsation threshold, or the spike rate in a single cell recording, can decrease when other sounds are added to the stimulus (Houtgast, 1973; Kiang et al., 1965). This effect is also called "two-tone suppression", "two-tone inhibition" or "lateral suppression" and is included in the more general terms "two-tone nonlinearity" and "two-tone interaction". The decrease in the loudness of the tone in question under these conditions is called "loudness reduction" (Houtgast, 1974a).

The results of our loudness matchings may be affected by such a loudness reduction. A reduction in the white-noise masking level when $\phi_1=\phi_{max}$ was never found. In fact, this level was always higher than in the absence of component I. In the masking-level procedure, the noise (i.e. the test signal) is presented simultaneously with the octave complex. Houtgast (1972) showed that simultaneous presentation of the stimulus and the test signal obscures the two-tone nonlinearity because the two are suppressed in the same way. In a similar way the masking noise cannot be used to detect some internal level of component II. This is a serious disadvantage of the masking-level procedure. We conclude that the method can be used to give a rough description of the phase effects, but that it is unsuitable for study of the contribution of the two-tone nonlinearity. The latter will be discussed in more detail in Chapter 6.

The day-to-day horizontal shift of the curves in Fig. 4.5.a and 4.5.b mentioned in section 4.2 explains the large variation in the magnitude of the phase effects (section 4.1). For instance, in Fig. 4.5.a for L_1=75 dB SPL the difference in masking level for $\phi_1=30^o$ and $\phi_1=120^o$ is 9 dB. This is the maximum phase effect for that (L_1,L_2) combination. If we suppose that the curves were shifted 2 dB to the left in another session, we would measure a maximum phase effect of 22 dB for the same (L_1,L_2) combination. The two cases are indicated by broken lines.

4.4. Conclusions

1) A distinct phase dependence of pitch, masking level and loudness of component II is found in octave complexes.
2) The magnitude of the effects depends on the levels of the components and on the frequency combination used.
3) The plot of pitch against phase resembles a sawtooth function. The maximum shift is 20% upwards.

4) The maximum phase effect in the loudness and masking level is 20 dB.

5) The phase condition under which the loudness or the masking level is minimum coincides with that for the jump in the pitch function.

6) The loudness data suggest that "loudness reduction" is operative.

CHAPTER 5

PITCH AND MASKED THRESHOLD IN OCTAVE COMPLEXES IN RELATION TO INTERACTION PHENOMENA IN TWO-TONE STIMULI IN GENERAL

5.1. Introduction

The common aspect of the experiments to be described in the present chapter is that a comparison is made of phenomena and functions in octave complexes with those in two-tone stimuli in which the frequency ratio of the tones is different from two. The experiments involved are pitch matchings, simultaneous-masking experiments and pulsation-threshold experiments. Because the aim of the pitch experiments is rather different from that of the masking and pulsation-threshold experiments, we introduce them separately.

5.1.1. *Introduction to the pitch experiments*

Octave complexes are not the only two-tone stimuli in which the pitch of one of the components can be influenced. Plomp (1967) found a phase-dependent pitch in a 200/600 Hz frequency combination. Walliser (1969) reported pitch shifts in a 4-kHz pure tone due to the presence of a second interfering tone. In that case there was no simple harmonic relation between the frequencies of the two tones. Each of them had a loudness level of 30 phones. The maximum pitch deviation corresponded to a positive frequency shift of 1.5%. Positive pitch shifts of a few percent were also found by Terhardt and Fastl (1971) for stimuli in which the frequencies of the tones had a ratio 2:3, 3:4 and 1:3. The tones had levels from 50 up to 70 dB SPL. The authors found no phase dependence of the pitch in these cases. The relation of the pitch effects in octave complexes to those in two-tone stimuli in general is the subject of the first part of this chapter.

Our conclusions here will also have implications for the theories of pitch perception. Until some years ago it was accepted by several authors that the pitch of a sound was determined by the reciprocal of the time interval between prominent peaks in the rectified waveform (see Chapter 2). Within the framework of that time theory, Plomp (1967) suggested that the sweep effect might be related to such a "time-interval detection". On the other hand the pitch effects in two-tone stimuli found by Walliser (1969) have never been considered in the light of the time theory, but in terms of a shift of a maximum in the masking pattern which was interpreted as an "excitation pattern". The latter idea originates from Egan and Meyer (1950). An extensive discussion of this explanation in spectral terms has been given by Terhardt (1972). Consideration of whether there is a relation between the pitch effects in all two-tone stimuli may help us to decide whether one and the same explanation will suffice to deal with all the different stimuli.

5.1.2. *Motivation of the masking experiments*

The masking experiments have been carried out to investigate the possibility that the masked threshold of a test tone one octave above the frequency of the masker might be affected by an aural second harmonic. A supporting argument frequently used in the literature (see Chapter 2) is that the curve showing the dependence of the simultaneously-masked threshold on L_1 has a slope of 2. In our experiments the simultaneously-masked threshold and the pulsation threshold for $f_t=2f_1$ was compared with that for test-tone frequencies below the octave.

5.1.3. *On- and off-ratio stimuli*

A stimulus in which the frequencies of the components have a ratio of m:n, where m and n are small integers will be called an "on-ratio stimulus". A stimulus with a frequency ratio between the components different from the above will be called an "off-ratio stimulus". This definition is rather flexible; for example, we will call the frequency combination (f_1,f_2) with $f_2=1.75f_1$ the off-ratio stimulus with respect to the on-ratio stimulus with $f_2=2f_1$.

5.2. Pitch experiments

5.2.1. *Pitch in off-ratio octave complexes*

The stimuli in the first pitch experiment were off-ratio octave complexes. The values of f_1 were 200 Hz, 400 Hz and 760 Hz; those of f_2 were $1.75f_1$, $1.90f_1$ (except for f_1=200 Hz) and $2f_1$−10 Hz. The procedure in these experiments is similar to that for the on-ratio octave complexes in section 4.2. We matched the pitch of a comparison tone with that of component II as a function of L_1 with L_2 as the parameter. The maximum value of L_2 was 45 dB SPL. L_1 was increased in steps of 1 or 2 dB over a range of 20 to 30 dB until component II was masked. A representative sample of our results is shown in Fig. 5.1. These three curves, all for the same value of f_1, have much the same shape. At lower values of L_1 the pitch of component II is often significantly higher than that of component II alone (indicated by the horizontal

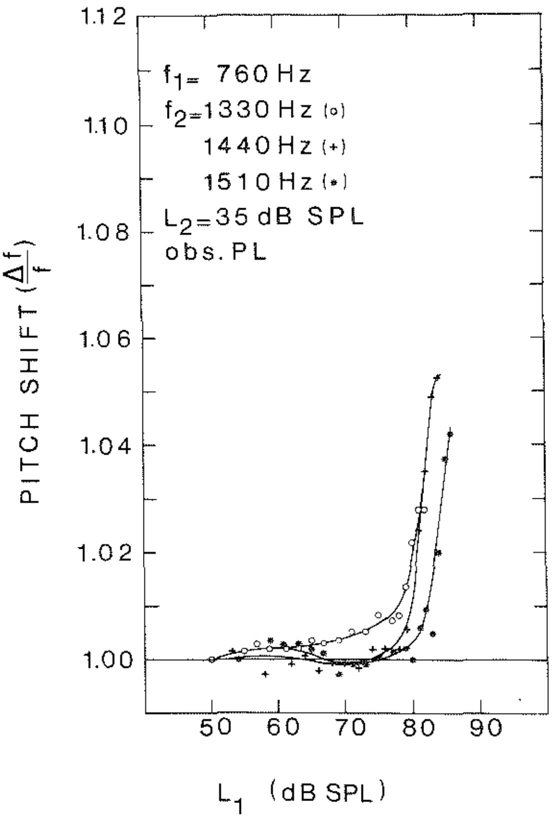


Fig. 5.1.

Pitch of component II as a function of L_1 for three frequencies f_2 in the region about one octave above f_1.

line). For f_2=1510 Hz the pitch first increases and then decreases to slightly below the original value. Finally, under all three conditions the pitch increases very sharply near the masked threshold of component II. At the same time the loudness decreases. Component II does not sound like a pure tone any more at these higher values of L_1: a change in timbre occurs. The maximum pitch effect does not depend on L_2. As regards the scatter in the data and the day-to-day variations, the remarks made in connection with the pitch and loudness experiments with on-ratio octave complexes in section 4.2. are applicable.

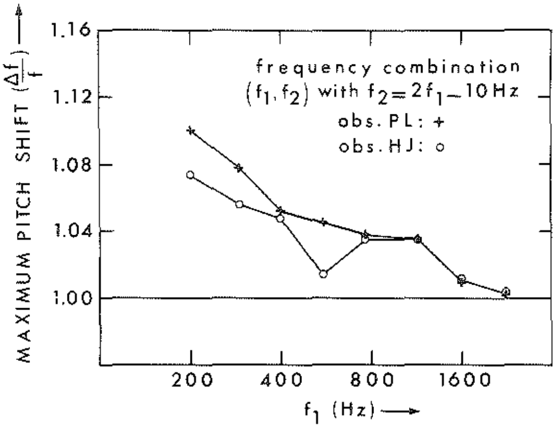
Comparison of the results obtained with on-ratio and off-ratio octave complexes reveals that the maximum pitch shift in an off-ratio octave complex is roughly half that in the corresponding on-ratio complex for $\phi_1=\phi_{min}$. This correlation is rather rough because of the spread in the highest experimental points in the figures.

5.2.2. *Maximum pitch shift in two-tone stimuli with $f_2=2f_1-10$ Hz for different values of f_1*

We determined the pitch of component II with $f_2 = 2f_1-10$ Hz as a function of L_1 for different values of f_1. The average of the highest two experimental points in the graphs (similar to those in Fig. 5.1) was defined as the maximum shift relative to f_2. We took the average of two experimental points to reduce the scatter. The maximum pitch defined in such a way is plotted in Fig. 5.2. as a function of f_1 for two observers.

Fig. 5.2.

Maximum pitch shift of component II of an off-ratio octave complex with $f_2 = 2f_1-10$ Hz as a function of f_1; L_2=35 dB SPL.



We saw in Chapter 4 that the maximum pitch shift in an on-ratio octave complex occurs when $\phi_1=\phi_{min}$. It was concluded in the preceding section that the maximum shift in an off-ratio complex is half that in the corresponding on-ratio complex for $\phi_1=\phi_{min}$. Therefore, apart from the factor two, Fig. 5.2. may be considered as giving the dependence of the maximum pitch effect in an octave complex on the frequenices of the components.

5.2.3. *Maximum pitch shift in two-tone stimuli with $f_2 = \frac{n}{m}f_1-10$ Hz as a function of $\frac{n}{m}$*

We measured the maximum pitch shift of component II with $f_2 = \frac{n}{m}f_1-10$ Hz as a function of $\frac{n}{m}$, where m and n are small integers and m<n. The procedure was as described in the previous section. The results for f_1=200 Hz are given in Fig. 5.3.a. and for f_2=1200 Hz in Fig. 5.3.b.

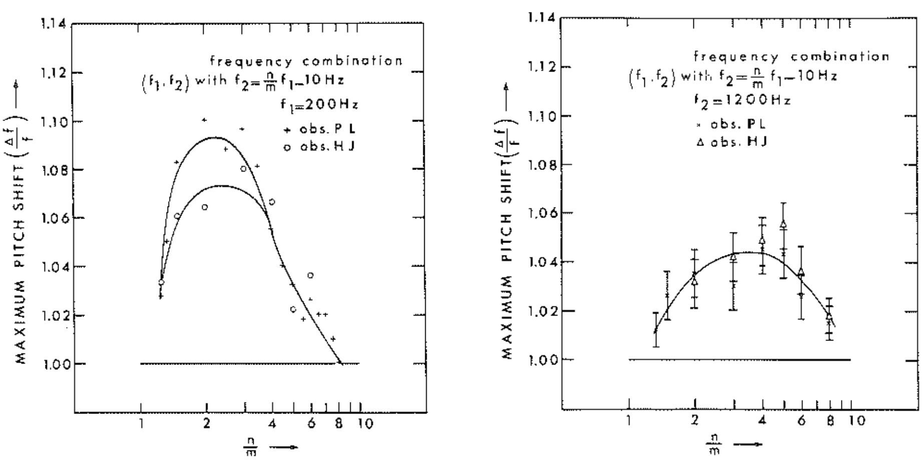


Fig. 5.3.

Maximum pitch shift of component II of a two-tone stimulus with $f_2 = \frac{n}{m}f_1-10$ Hz as a function of $\frac{n}{m}$, m and n being small integers and m<n, L_2=35 dB SPL.
a) f_1 is constant (200 Hz)
b) f_2 is constant (1200 Hz)

We concluded in section 5.2.2 that the maximum pitch effect in a stimulus with $f_2 = 2f_1-10$ Hz is a measure of that in the corresponding on-ratio octave complex. We extrapolated this relation to the frequency ratios m:n used in the present section. It was checked for two ratios (1:3 and 2:3) that the maximum pitch shift in the on-ratio situation is twice that in the corresponding off-ratio situation, just as with octave complexes. We may, therefore, conclude that, apart from the factor two, Fig. 5.3. holds for the corresponding on-ratio stimuli too.

5.2.4. *Discussion*

The experimental results indicate that pitch effects are not exclusive to octave complexes. Moreover, we found a correlation between the pitch effects in off-ratio octave complexes (Fig. 5.1) and those in on-ratio complexes. Pitch effects are also found for other frequency ratios. The magnitude of the pitch effects for the different ratios, plotted against $\frac{n}{m}$ can be fitted by a single curve (Fig. 5.3). This also holds for more complex ratios such as 2:7; it would thus seem as if the shifts in the pitch of the higher-frequency component in all two-tone stimuli represent the same phenomenon.

Under the level conditions chosen, the magnitude of the pitch shift depends on three factors. First, on the value of L_1 relative to that of L_2 (we may note here that the shape of the curves of Fig. 5.1 does not depend on L_2). Secondly, on whether the frequencies of the components are relatively high or low. Fig. 5.2. illustrates this for octave complexes. The third factor determining the size of the pitch shift is the frequency ratio of the components. The largest effects are found for the ratios 1:2 to 1:3. The difference in height between the curves of Fig. 5.3.a and 5.3.b must be due to the second factor.

The relation between the phase-dependent on-ratio and off-ratio pitch effects is such that the maximum effect in the off-ratio situation is equal to the maximum effect for the on-ratio stimulus averaged over the phase conditions.

The experiments described so far are closely related to investigations of the effect on the pitch of a tone due to the simultaneous presentation of a white noise, a band noise or a low-pass noise, Such investigations were initiated by Egan and Meyer (1950) and continued by various authors (see e.g. Terhardt (1972) and Van den Brink (1975). The results of all these experiments with interfering sounds, (either a noise or a tone) show four common features: pitch shifts are usually only due to sounds with a frequency content lower than that of the stimulus frequency; pitch shifts are mainly upwards; the pitch shifts increase with increasing level of the interfering sound; and pitch effects only occur if the tone is partially masked, so that its loudness is reduced. Our results are generally in agreement with these findings. However, a decrease in loudness is not always accompanied by an increase in pitch as may be seen in Fig. 4.4 for $\phi_1=20^o$.

5.3. Pure-tone masking experiments

5.3.1. *Pulsation-threshold method*

The stimulus frequencies (f_1) were 200 Hz, 400 Hz, 760 Hz and 1600 Hz, and those of the test tones $1.75f_1$, $1.90f_1$ except at f_1=200 Hz) and $2f_1$. We measured the pulsation threshold of the test tones as a function of L_1 in the range from 60 to 90 dB SPL. A sample of the results is shown in Fig. 5.4. For a given masker, the curves have much the same shape for all test-tone frequencies. The scatter in the experimental points is 3 dB.

To classify the increase in the pulsation threshold relative to L_1, the measured curves were divided into two

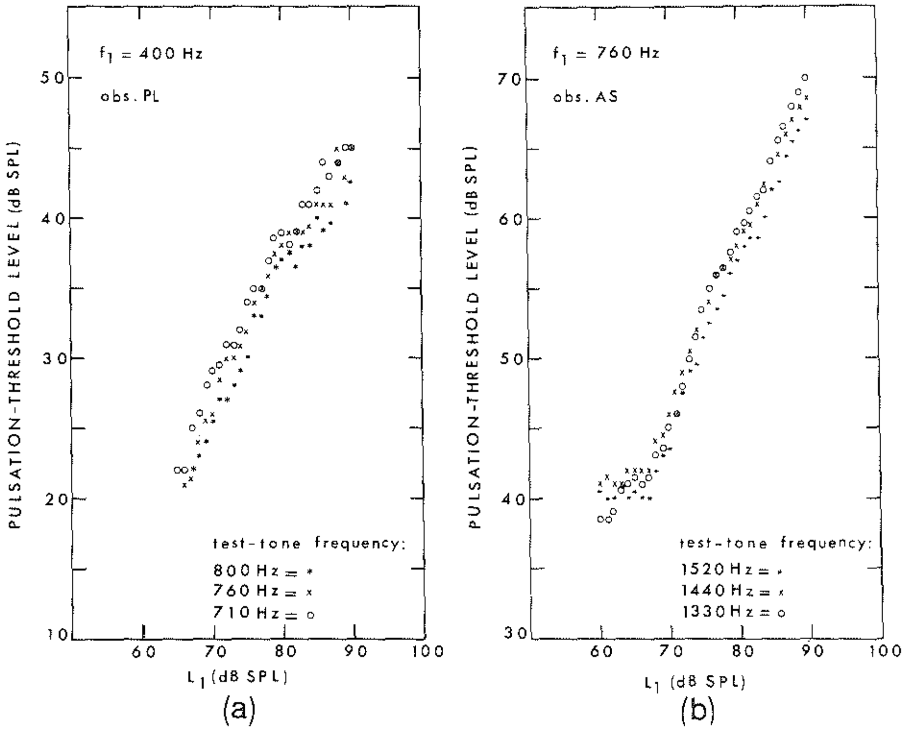


Fig. 5.4.

Pulsation threshold as a function of L_1 for two values of f_1 and two observers. Test tones in the frequency region about one octave above f_1.
a) f_1=400 Hz, observer PL
b) f_1=760 Hz, observer AS

parts, one for L_1 below 75 dB SPL and one above that. The relative increase in each of those parts, expressed as decibels *vs.* decibels, is given in Table 5.I for some values of f_1. The scatter in these values is 0.2.

obs.	f_1 = 200 Hz		f_1 = 400 Hz		f_1 = 760 Hz		f_1 = 1600 Hz	
PL	1	1	1.4	1	1.4	1	decr	1
AS	1	1	1.4	1	1.4	1	decr	1
RE	1	1	1.4	1	1.4	1	decr	1
WJ	1	1	decr	1	decr	2.4	decr	1

Table 5.I

The increase of the pulsation threshold for pure-tone stimuli relative to L_1, expressed as decibels *vs*, decibels, for different stimulus frequencies f_1 and $f_t=2f_1$. The left-hand half of each column corresponds to the parts of the curves of Fig. 5.4. for L_1 below 75 dB SPL, and the right-hand half to the parts above 75 dB SPL. The indication "decr" means that the increase gradually diminishes with decreasing L_1 in the part of the curve in question.

5.3.2. *Simultaneous masking*

The stimuli and test-tone frequencies used here were the same as in the preceding section. The highest test-tone frequency was $2f_1$-10 Hz, in order to avoid slow beats. A sample of the results is shown in Fig. 5.5. The scatter in the experimental points is about 3 dB. The relative increase of these functions, expressed as decibels *vs*. decibels, is given in the upper two rows of Table 5.II. The experimental error in these values is again 0.2.

These experiments were also performed for test-tone frequencies of exactly $2f_1$, with $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$ which correspond to the minimum and maximum masked threshold respectively. The relative increases under these two phase conditions are given in the bottom two rows of Table 5.II.

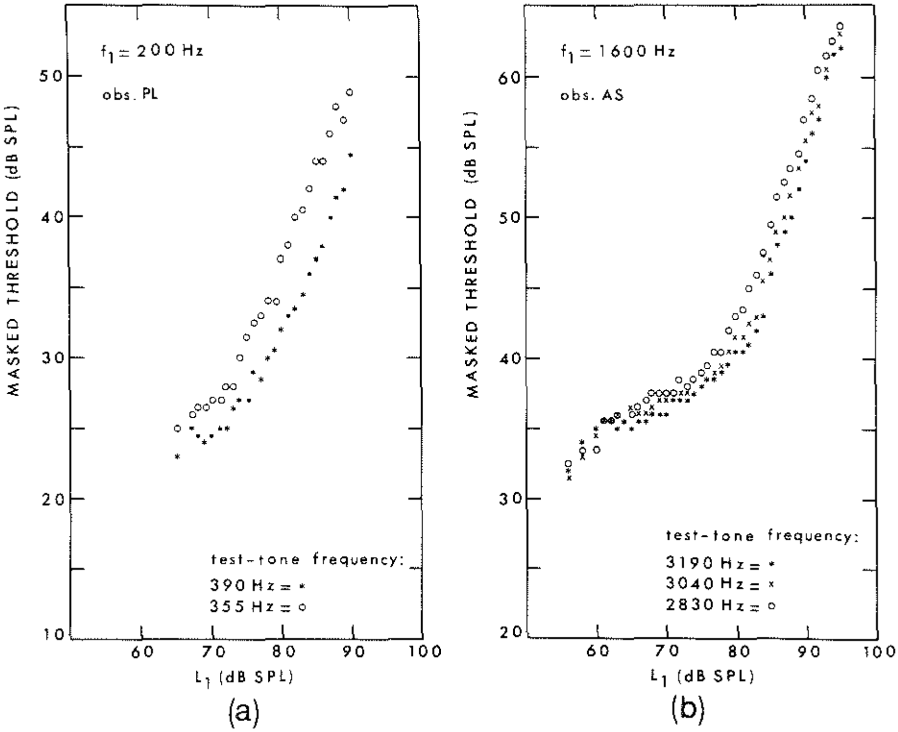


Fig. 5.5.

Simultaneously-masked threshold as a function of L_1 for two values of f_1 and two observers. Test tones in the frequency region around $2f_1$.

obs.	f_1 = 200 Hz	f_1 = 400 Hz	f_1 = 760 Hz	f_1 = 1600 Hz
PL	1.2	1.8	2.4	2.6
AS	1.1	2.2	2.4/2.0	1.6
PL (oct) -	1.0	2.2	2.3	
PL (oct) +	1.0	2.0	2.0	

Table 5.II.

Relative increase in the simultaneously-masked threshold of component II with respect to L_1, expressed in decibels *vs.* decibels, for different values of f_1. The data in the top two rows correspond to conditions under which the frequency of the test tone is less than $2f_1$. The bottom two rows correspond to conditions under which the frequency of the test tone is exactly $2f_1$. The plus and minus signs indicate the phase relations between masker and test tone which yield the maximum and minimum masked threshold respectively. The two values for observer AS at f_1=760 Hz indicate two different relative increases in the curve, one for L_1<80 dB SPL (2.4) and the other for L_1>80 SPL (2.0).

5.3.3. *Discussion*

The similarity between the curves for the different test-tone frequencies with a constant value of f_1 is striking. The data of Table 5.II also indicate that the relative increase in the masked threshold for $f_t=2f_1$ hardly differs from that for f_t less than $2f_1$. The relative increase for $f_t=2f_1$ agrees reasonable well with data given in the literature by Clack et al. (1972) and Nelson and Bilger (1974).

Over a larger test-tone frequency range, the relative increase in the functions can vary, as shown by the results of the pure-tone masking experiments performed e.g. by Wegel and Lane (1924), Egan and Hake (1950) and Nelson and Bilger (1974). The slopes of their curves representing the masked threshold of the test tone as a function of the level of the masker vary when the frequency of the test tone is increased from $1.5f_1$ to $3f_1$.

The implications of the results for a possible explanation of the phase effects in terms of aural harmonics will be discussed in Chapter 8.

5.4. Conclusions

1) The pitch effects in octave complexes described in Chapter 4 are related to pitch effects found in two-tone stimuli in general under the same level conditions.
2) The magnitude of the pitch shift in a two-tone stimulus depends on the value of L_1 with respect to that of L_2, on the absolute frequencies of the components and on the ratio of f_2 to f_1.
3) The pitch shift as a function of the frequency ratio of the components is maximum for ratios between 1:2 and 1:3.
4) The combination 1500/3000 Hz represents the upper limit for pitch effects in octave complexes.
5) Pitch effects larger than 1 to 2% occur only under conditions of partial masking. Partial masking, however, does not always yield pitch effects.
6) Simultaneous-masking experiments and pulsation-threshold experiments with pure-tone stimuli (frequency f_1) reveal no anomalous behaviour for $f_t=2f_1$ as compared with the results for values of f_t less than $2f_1$.

PART III

PULSATION-THRESHOLD EXPERIMENTS WITH OCTAVE COMPLEXES

CHAPTER 6

TWO-TONE INTERACTION IN OCTAVE COMPLEXES

6.1. Introduction

The data presented in Chapter 4 revealed distinct phase effects in the pitch and the loudness of component II. The experiments to be described in this chapter arose from questions about possible mechanisms underlying the phase effects, such as: Is there a shift in the peak of the masking pattern of the complex related to the pitch shift as a function of phase? In which "frequency channel" of the auditory system do the phase effects come about? If there is an internal representation of the stimuli in a specific frequency region, how do the phase effects in it depend on the levels of the components?

These experiments have been performed by the pulsation-threshold method, for the reasons explained in Chapter 3. The involvement of "two-tone suppression" enables us to compare the results with the loudness data presented in Chapter 4 where the phenomenon of loudness reduction was found. The experiments were carried out with both on- and off-ratio complexes.

6.2. Experiments

6.2.1. *Pulsation-threshold patterns of octave complexes*

Pulsation-threshold patterns, representing the pulsation threshold as a function of the test-tone frequency, have been measured for the octave complexes 200/400 Hz, 400/800 Hz and 760/1520 Hz. Typical results are shown in Fig. 6.1. The values of L_1 and L_2 correspond to conditions under which the loudness and pitch of component II are affected by the phase, as used in Chapter 4. The measurements were performed for test-tone frequencies around f_2 under the phase conditions corresponding to maximum $\phi_1=\phi_{max}$ and minimum audibility $\phi_1=\phi_{min}$ of component II. The patterns for the separate components were measured in the same session, and are indicated schematically in Fig. 6.1.

The results reveal differences between the pulsation thresholds for $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$ in the test-tone frequency range from 0.80 f_2 to 1.25 f_2. A representative value of the threshold difference for $f_t=f_2$ is 10 dB. The difference decreases with increasing frequencies of the components. The peak in the pattern of the complex around $f_t=f_2$ appears to be invariably broader than the pattern for component II alone, indicated by the broken lines.

The adjustments for the different test-tone frequencies were made in a random order. Their scatter is 3 dB. There was a session-to-session variation in the difference between the pulsation thresholds for $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$ for a given level combination and test-tone frequency. The difference found in a previous session could be reproduced in the next session by changing the value of L_1 by a few decibels.

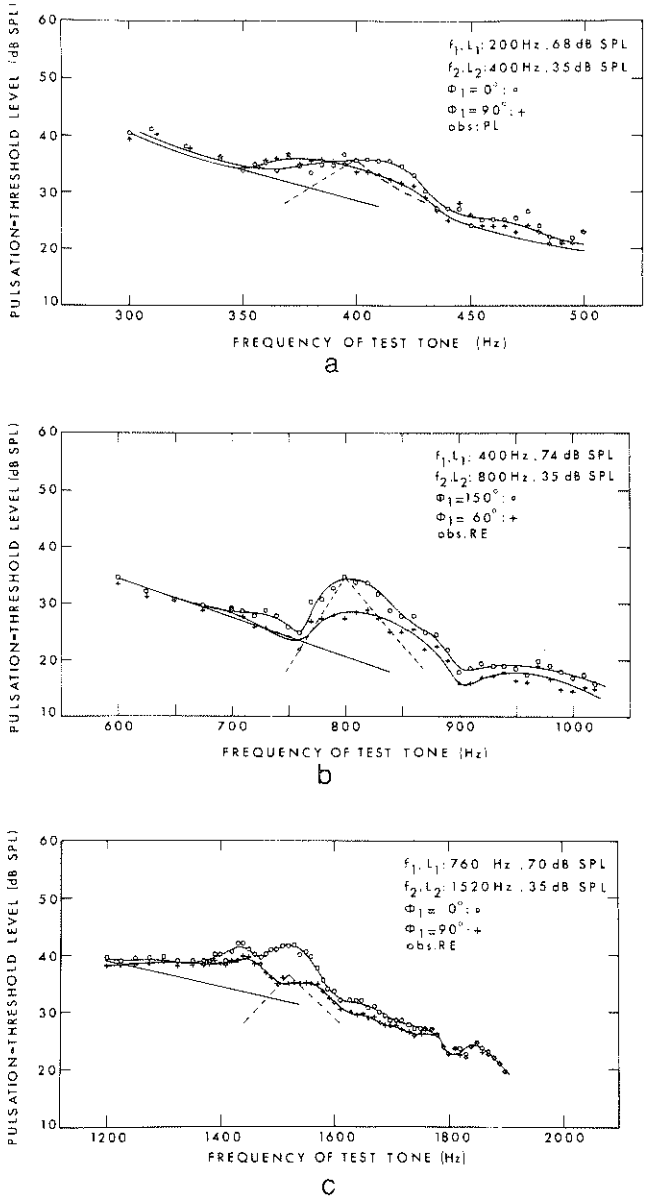


Fig. 6.1.

Pulsation-threshold patterns of octave complexes. The experimental points indicated by circles correspond to the phase condition $\phi_1=\phi_{max}$, and those represented by plusses to $\phi_1=\phi_{min}$. The straight lines indicate the pulsation pattern of the lower-frequency component alone, while the broken lines indicate the pattern for the higher-frequency component alone.

6.2.2. *Pulsation threshold as a function of L_1 for on-ratio octave complexes with $f_t=f_2$*

The next step was to vary L_1, with f_t equal to f_2. The pulsation threshold was measured with L_2 as the parameter for $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$. The corresponding level for component I alone was also measured. The results for three values of L_2 are plotted in Fig. 6.2. The curves for the different values of L_2 were measured on different days. They were shifted in a horizontal direction by a few decibels until all curves came together at the highest values of L_1 used.

The pulsation threshold for the complex is equal to that for component II alone (horizontal full lines) at relatively low values of L_1, and is equal to that for component I alone at relatively high values of L_1 (the dashed curve). The pulsation threshold in the intermediate range is phase-dependent. When $\phi_1=\phi_{min}$, it first decreases with increasing value of L_1 until it is lower than that for component I alone. As L_1 increases further, the pulsation threshold remains below that for component I but approaches closer and closer to the latter.

The shape of the curves for $\phi_1=\phi_{max}$ depends strongly on the value of L_2. There is a pronounced drop below the level for component II alone for L_2=45 dB SPL. The opposite effect is found for L_2=25 dB SPL. In this case the pulsation threshold for the complex at L_1=72 dB SPL is 8 dB above that for each of the components, while the increase of the pulsation threshold relative to L_1 and L_2, expressed in decibels *vs.* decibels, is 1 for component II and 1.6 for component I (for L_1=72 dB SPL).

A set of observations similar to those of Fig. 6.2. was also made by observer RE for the frequency combination 760/1520 Hz. His results showed the same trend, as did the individual results of two other observers.

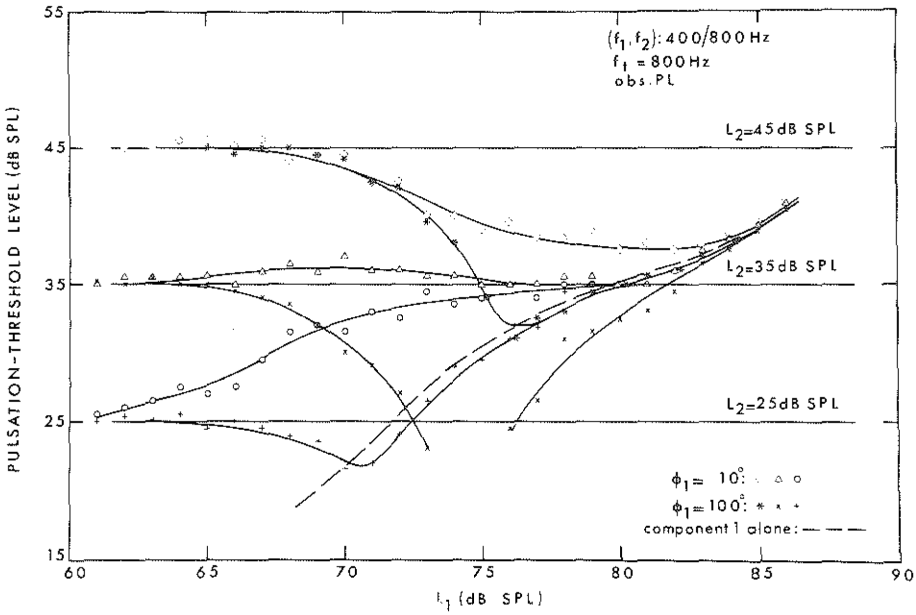


Fig. 6.2.

Pulsation threshold for an octave complex as a function of L_1 for three different values of L_2. Frequency combination 400/800 Hz. Test-tone frequency 800 Hz. The symbols ◇, □ and ○ correspond to $\phi_1=\phi_{1,max}$ (10°) and the symbols * × + to $\phi_1=\phi_{1,min}$ (100°). The horizontal lines correspond to the pulsation threshold for the lower-frequency component alone, measured as a function of L_1. The absence of experimental points in the curve for L_2=35 dB SPL and $\phi_1=\phi_{1,min}$ between L_1=73 and 76 dB SPL indicates that no adjustment could be made due to the low level of the test tone. It does not necessarily mean that there really is a gap in the curve.

The experimental points show a scatter of 3 dB. Systematic variation of the data from session to session was also found. This is more pronounced at higher values of L_1 (up to 5 dB). The shape of the curves does not change much but the curves as a whole are shifted in a horizontal direction by a few decibels.

6.2.3. *Relation between pulsation thresholds for off- and on-ratio stimuli*

The results of the experiments with octave complexes described in the previous section show that the pulsation threshold for the complex is lower than that for component II alone, e.g. for the level combination 78/45 dB SPL with $\phi_1=\phi_{max}$ and $\phi_1=\phi_{min}$, and for the level combination 73/25 dB SPL with $\phi_1=\phi_{min}$. Can these reductions be identified with the phenomenon of "two-tone suppression" demonstrated by the psychophysical experiments of Houtgast (1972)? In answering this question we thought it better to start not from results obtained for on-ratio octave complexes but from comparable experiments with two-tone stimuli in general, without the choice of a phase relation (and possible internally generated components) being involved.

A similar question can be asked in connection with the finding that the pulsation threshold for the complex is enhanced by 8 dB with respect to that for the separate components at (L_1,L_2)= 72/25 dB SPL and $\phi_1=\phi_{max}$. Is this "enhancement" a general two-tone effect or is it a consequence of the specific stimulus?

We designed two experiments to answer these questions; they were performed by a single observer. In the first experiment the stimulus was the off-ratio octave complex 760/1330 Hz. We measured the pulsation threshold for f_t=1330 Hz as a function of L_2 for two different values of L_1 (unlike section 6.2.2 where we measured the pulsation threshold as a function of L_1 with L_2 as the parameter). We chose the two values of L_1 on the basis of the results of Fig. 6.2: a relatively low value, where enhancement could be expected and a relatively high one at which reduction in the pulsation threshold could be expected. The adjustments were made over the whole range of values of L_2 for which interaction could be found at a given value of L_1. The results are shown in Fig. 6.3.a.

The same experiment was repeated with the on-ratio frequency combination 760/1520 Hz, a test-tone frequency of 1520 Hz and $\phi_1=\phi_{max}$. The results shown in Fig. 6.3.b. are very similar to those of Fig. 6.3.a. This similarity implies that the results found for octave complexes under the phase condition giving maximum audibility of component II ($\phi_1=\phi_{max}$) are representative of those for two-tone effects in general.

Fig. 6.3.a. shows that the pulsation threshold is lower than that for component II alone only for L_1=87 dB SPL and L_2 beyond 45 dB SPL. The decrease in pulsation threshold for octave complexes at higher values of L_1 and L_2 e.g. (L_1, L_2) = 78/45 dB SPL (Fig. 6.2) and 87/55 dB SPL (Fig. 6.3.b) may thus be identified with two-tone suppression.

"Enhancement", as found e.g. in Fig. 6.2. for L_2=25 dB SPL and $\phi_1=\phi_{max}$, is much more difficult to define because we do not know how large an increase in pulsation threshold to expect if two non-interacting tones which give rise to the same pulsation threshold are added. Naturally this increase will depend on the relative increase for each component separately. Taking into account that the relative increase in the pulsation threshold for the separate components expressed in decibels *vs* decibels does not differ much from 1, we may regard a 3-dB increase as "normal" if the tones have no fixed phase relation and interaction is absent. The increase actually found, however, at lower values of L_1 and L_2 (72/30 dB SPL in Fig. 6.3.b. and 71/25 dB SPL in Fig. 6.2) was always 6 to 8 dB. Because this effect is so consistently found in our results, we propose the term "two-tone enhancement" for it. However, further investigation is needed to explore the nature of this phenomenon.

An effect found only in on-ratio octave complexes is

Fig. 6.3.a.

Pulsation threshold as a function of L_2 for two different values of L_1. Frequency combination 760/1330 Hz. Test-tone frequency 1330 Hz. The horizontal lines correspond to the pulsation threshold for component I alone at the levels indicated. The oblique straight line of slope 1 corresponds to the pulsation threshold for the higher-frequency component alone as a function of L_2.

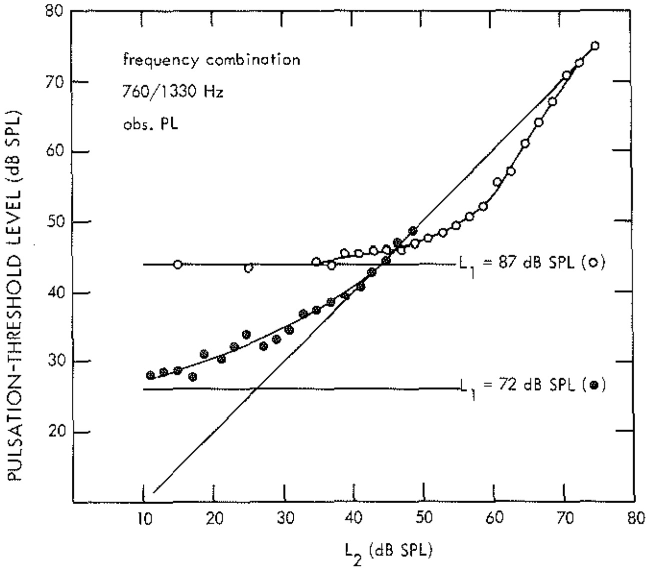
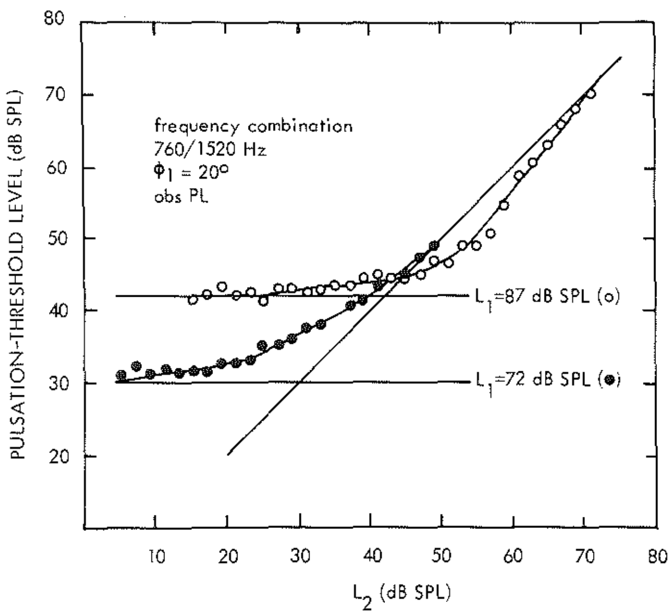


Fig. 6.3.b.

As Fig. 6.3.a , for the frequency combination 760/1520 Hz, test-tone frequency 1520 Hz and $\phi_1=\phi_{max}$.



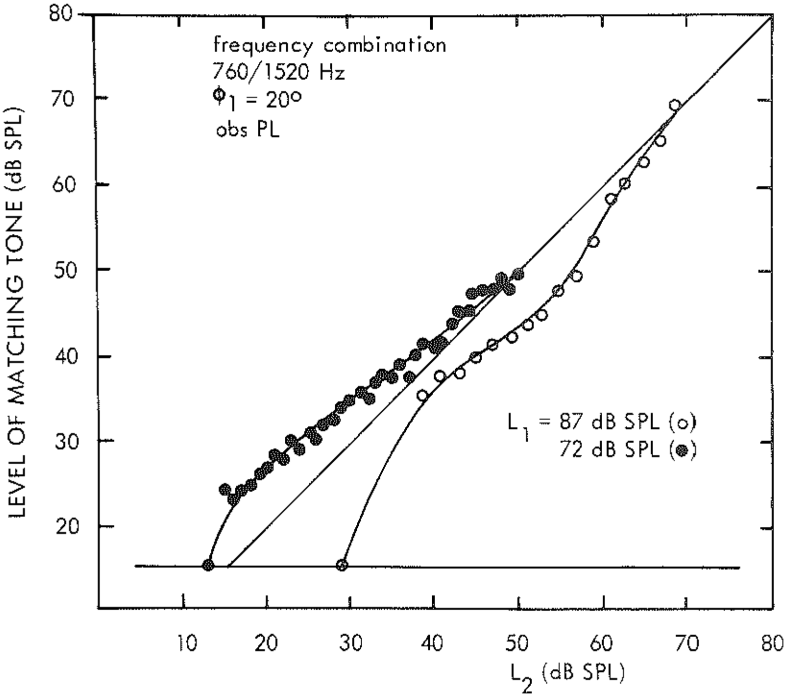


Fig. 6.4.

Loudness of component II as a function of L_2 for L_1=72 dB SPL and 87 dB SPL. Frequency combination 760/1520 Hz. $\phi_1 = \phi_{max}$. The horizontal line indicates the absolute threshold of the matching tone. The oblique line of slope 1 indicates the loudness of component II of the complex alone, as a function of L_2.

the decrease in the pulsation threshold to below that for the lower-frequency component alone. This effect is not called "suppression" as it could have another origin.

6.2.4. *Relation between pulsation threshold and loudness*

The data presented in Chapter 4 showed that the loudness of component II in an octave complex was often lower than that of component II alone for $\phi_1=\phi_{max}$. In the preceeding sections we found that the pulsation threshold for the complex, even at maximum audibility, was below that for component II alone for some level conditions, e.g. 78/45 dB SPL. Is this effect associated with loudness reduction for these level combinations? On the other hand, the pulsation threshold could be enhanced,e.g. for the combination 72/25 dB SPL. Is the loudness enhanced as well in this case? To check this, we matched the loudness of component II for the same stimulus and under the same conditions as used in the pulsation-threshold experiment described in the previous section (frequency combination 760/1520 Hz). The results are shown in Fig. 6.4.

Comparison of Fig. 6.3.b. and 6.4. shows that a reduction in the loudness of component II corresponds to the presence of suppression in the pulsation-threshold results. Furthermore, enhancement in the pulsation-threshold results corresponds to a distinct enhancement of the loudness of component II.

6.3. Discussion

6.3.1. *Pulsation-threshold patterns*

The results of section 6.2.1 show that phase effects occur in a frequency region around f_2. This indicates that the region of interaction is roughly the same as that stimulated by component II itself under the level conditions chosen in our experiments.

The question now arises whether the peak in the pulsation pattern of the complex around $f_t=f_2$ shifts with phase, in the same way as the pitch does. We, therefore, measured pulsation-threshold patterns under phase conditions between the two extreme ones, but found no systematic shift of the peak as a function of phase for the combinations 400/800 Hz and 760/1520 Hz. Whether there is a shift for the 200/400 Hz frequency combination is less easy to determine, as the patterns for the different phase conditions are rather flat. So far, therefore, we have no firm evidence for a shift of the peak as a function of the phase. However, we have found that the phase at which the pulsation threshold is minimum for a given test-tone frequency depends on that frequency. The consequences of such a dependence will be dealt with in the next chapter.

6.3.2. *Pulsation threshold and loudness*

It may be concluded that the results obtained by the pulsation-threshold method and by loudness matching have much in common. This supports the view that loudness reduction is caused by "two-tone suppression". On the other hand "loudness enhancement", as far as we know, is a new effect which needs further study.

A more quantitative comparison of the results of the two methods has been made for the 760/1520 Hz frequency combination with reference to the data of Figs. 6.3.b and 6.4. The loudness of component II was plotted against the pulsation threshold for the same level conditions. The results are shown in Fig. 6.5. (right-hand and left-hand curves). The broken curve is from a different type of experiment, in which both the loudness and the pulsation threshold were estimated as a function of phase for a given level combination. The loudness and the pulsation threshold under each of these phase conditions is represented as a dot in the figure. It can be seen that the results of this experiment fit in well with the other data.

Fig. 6.5. shows that the curves are very steep, when the

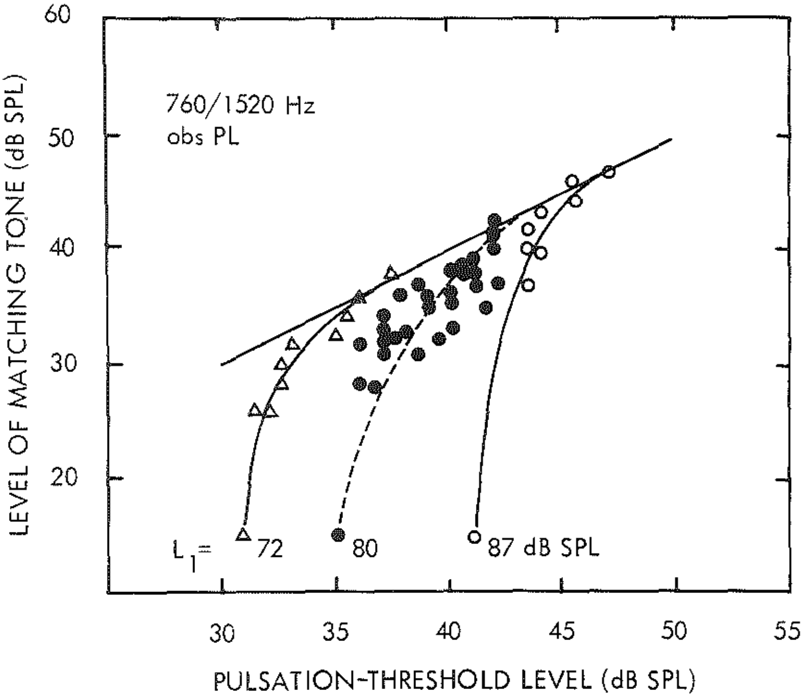


Fig. 6.5.

Loudness of component II of an octave complex as a function of pulsation threshold for a test-tone frequency $f_t=f_2$ and three values of L_1. Frequency combination 760/1520 Hz. The lowest point on each curve corresponds to the absolute threshold of the matching tone. The oblique line represents the relation between loudness and pulsation threshold when the lower-frequency component is absent.

pulsation-threshold for the complex is not much higher than that for component I alone. The curve for the highest value of L_1 shows the greatest slope. The curves approach a relative increase of 1 as the pulsation threshold is raised more and more. The results of Fig. 6.5 illustrate that relatively small variations in pulsation threshold due to the phase gives rise to as large effects on loudness (up to 20 dB) as those reported in Chapter 4.

6.4. Conclusions

1) Pulsation-threshold experiments performed with octave complexes at $f_t=f_2=2f_1$ reveal phase effects.
2) The test-tone frequency region in which the pulsation threshold depends on phase extends from $0.80f_2$ to $1.25f_2$ for $L_2<45$ dB SPL; This region is about the same as that stimulated by component II alone. There is no systematic shift of the component II peaks in the pulsation pattern of the complex as a function of the phase.
3) The maximum effect of phase on the pulsation threshold is found for $f_t=f_2$, and varies between 7 and 12 dB. This value depends on the experimental procedure and decreases with increasing frequency.
4) The pulsation threshold for octave complexes exhibits two-tone suppression at higher values of L_1 and L_2, e.g. for the combination 78/45 dB SPL. The reverse phenomenon, i.e. an increase in pulsation threshold for the complex by 6 to 8 dB when those for the separate components are equal, is found at lower values of L_1 and L_2, e.g. for the combination 72/25 dB SPL. We call this phenomenon "enhancement". It seems to be a general two-tone phenomenon, as it is also encountered in the results for off-ratio octave complexes with $f_2=1.75f_1$ under the same level conditions. An effect which is exclusive to on-ratio octave complexes is the decrease in the pulsation threshold to below that for component II itself for some values of L_1, L_2, and ϕ_1.
5) The loudness of the higher-frequency component of an octave complex is closely related to the pulsation threshold for the complex at $f_t=f_2$. Suppression in the latter case corresponds to reduction of the loudness, while enhancement of the pulsation threshold corresponds to a rise in loudness.

CHAPTER 7

FREQUENCY DEPENDENCE OF PHASE RELATIONS IN TWO-TONE OCTAVE COMPLEXES

7.1. Introduction

The first part of the present chapter deals with the question of whether the phase at which the pulsation threshold is maximum or minimum for a given frequency combination depends on the test-tone frequency. The existence of such phase-frequency relation might have implications for the phase-pitch relation found in Chapter 4.

The second part deals with the question of whether the phase at which the loudness of component II is minimum or maximum corresponds to the same waveform of the movement of the basilar membrane for all frequency combinations. To answer this question we need to know the phase characteristic of the outer/middle-ear system in addition to the results of the experiments mentioned above. The result of a determination of this phase characteristic is also presented.

7.2. Phase for minimum and maximum pulsation threshold as a function of f_t for different frequency combinations

7.2.1. *Experiments and results*

We determined cross-sections of the pulsation patterns of octave complexes shown in Chapter 6 by measuring the pulsation threshold level as a function of the phase with

f_t as parameter. The results for the frequency combination 200/400 Hz for three values of f_t are shown in Fig. 7.1. We see that the phase at which the threshold is minimum and maximum depends on f_t. The phase at which the minimum occurs was then measured systematically as a function of f_t for the frequency combinations 200/400 Hz, 400/800 Hz and 760/1520 Hz. The results are shown in Fig. 7.2. Each experimental point was obtained by decreasing the value of L_t until the test tone sounded continuously under nearly all phase conditions except in a "dip" where it was still pulsating. Similar results (not shown here) were obtained with observer WJ.

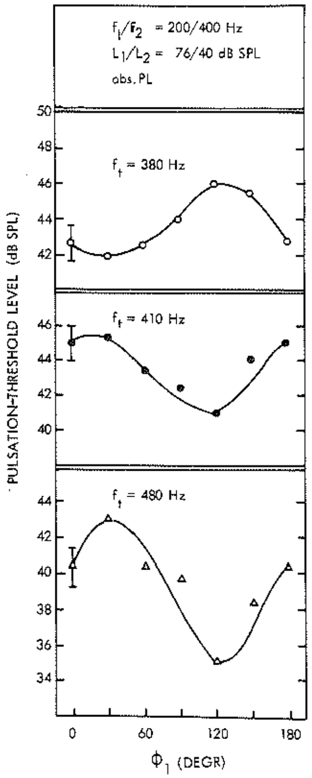


Fig. 7.1.

Pulsation threshold as a function of ϕ_1 for three different values of f_t. Frequency combination 200/400 Hz.

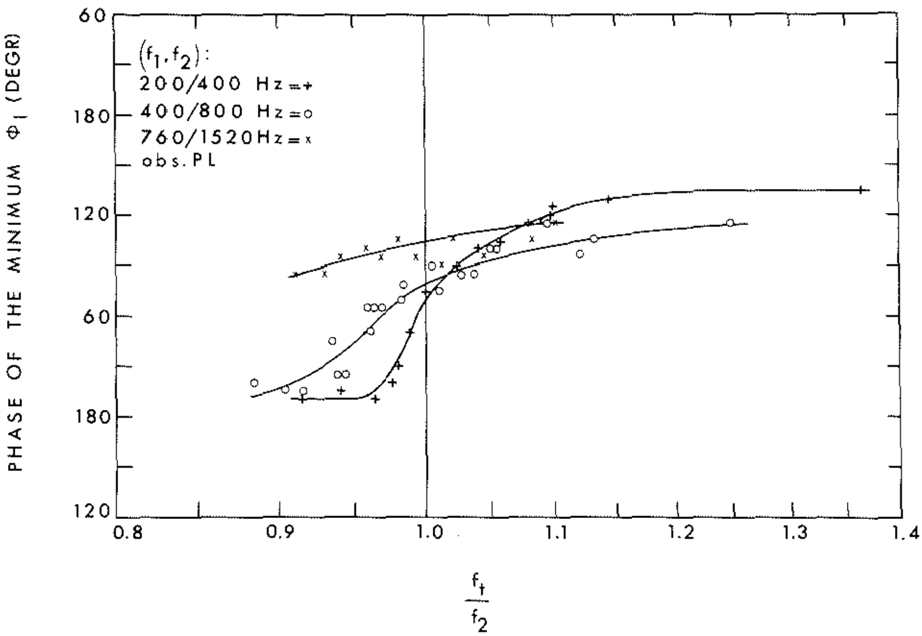


Fig. 7.2.
Phase (ϕ_1) at which the pulsation threshold is minimum, as a function of the ratio of f_t and f_2 for three frequency combinations.

The phase at which the pulsation threshold is minimum was found not to depend on the values of L_1 or L_2 (for L_2 in the range up to 45 dB SPL) for a given value of f_t.

The beginning and end of the curves of Fig. 7.2. correspond to the extreme value of f_t at which phase effects could be observed. The upper limiting value shifted slightly to higher frequencies at higher values of L_1 and L_2. The curves in Fig. 7.2. represent the maximum range for L_1<45 dB SPL.

In another experiment we investigated how the range of values of f_t for which phase effects can be observed depends on the frequency combination. The size of this interaction region is expressed in terms of the ratio of the maximum and

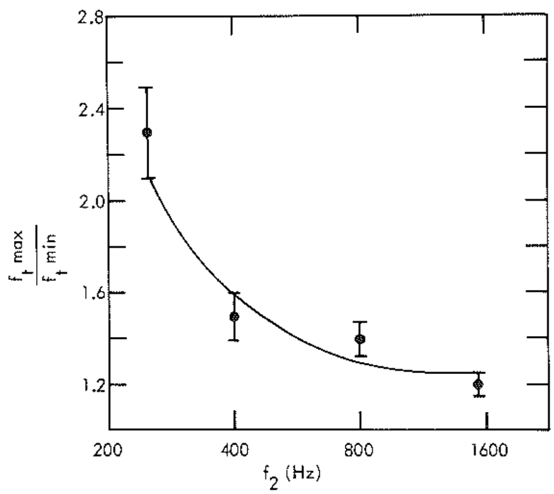


Fig. 7.3.

Range of values of f_t for which phase effects can be found in the pulsation threshold for octave complexes, for four values of f_2. The range is expressed as the ratio of the maximum and minimum values of f_t at which phase effects can be found. The curve represents the dependence on frequency of a bandwidth which is a constant factor of 1.25 for frequencies beyond 1000 Hz and a constant amount of 250 Hz below 1000 Hz.

the minimum values of f_t at which phase effects can be found. This ratio is plotted in Fig. 7.3. as a function of f_2.

7.2.2. *Discussion*

Comparing Fig. 7.2. with Fig. 4.1, we may conclude that for the frequency combinations 400/800 Hz and 760/1520 Hz the values of ϕ_1 at which component II is maximally and minimally audible coincide with the phases of maximum and minimum pulsation threshold, respectively, for $f_t=f_2$. For the 200/400 Hz combination, however, the value of ϕ_{min} (90^o) is somewhat higher than the phase for minimum pulsation threshold at $f_t=f_2$ ($\phi_1=70^o$). Since this difference is so small, we shall continue to use ϕ_{max} and ϕ_{min} to define the extremes of both the loudness and the pulsation threshold.

The curve of Fig. 7.3. represents a critical-band-like dependence on frequency: a constant factor above 1000 Hz and a constant amount for lower frequencies. The curve was drawn so as to pass through the experimental point for f_2=1520 Hz; it may be seen that the other three experimental points then give a reasonable fit. Although the width of the interaction region for phase effects is nearly a factor two larger than the width of the critical band (250 Hz as compared with 150 Hz at 1000 Hz), the similar dependence on frequency in the two cases suggests a relation.

Fig. 7.2. gives only the phase conditions corresponding to the minimum pulsation threshold. Fig. 7.1. shows that the phases for the maxima differ 90° from those for the minima. It thus is plausible that the dependence of the phase for maximum pulsation threshold on f_t will be given by the curves of Fig. 7.2. as well, if they are shifted 90° upwards or downwards.

7.2.3. *Implications for the pitch effects*

As we have just seen, there is a distinct dependence of the phases for minimum and maximum pulsation threshold on f_t for the 200/400 Hz combination. Now what implications does this result have for the pitch effects? Fig. 7.4. gives an idealized plot of the phases for minimum and maximum pulsation threshold as functions of f_t for the combination 200/400 Hz, as derived from Fig. 7.2.* From now on we restrict ourselves to the consideration of the maxima. From Fig. 7.4. we can derive how the test-tone frequency at which the pulsation threshold is in a maximum depends on ϕ_1. Expressing this value of f_t as a shift relative to f_2 gives the graph shown in the left-hand part of Fig. 7.5. The same procedure applied to the frequency combination 400/800 Hz gives the middle part of Fig. 7.5. and applied to the 760/1520 Hz combination it gives the right-hand part of Fig. 7.5.

*We give an idealized graph here as we are particularly interested in the maxima. These were not measured directly, as they were more difficult to estimate than the minima.

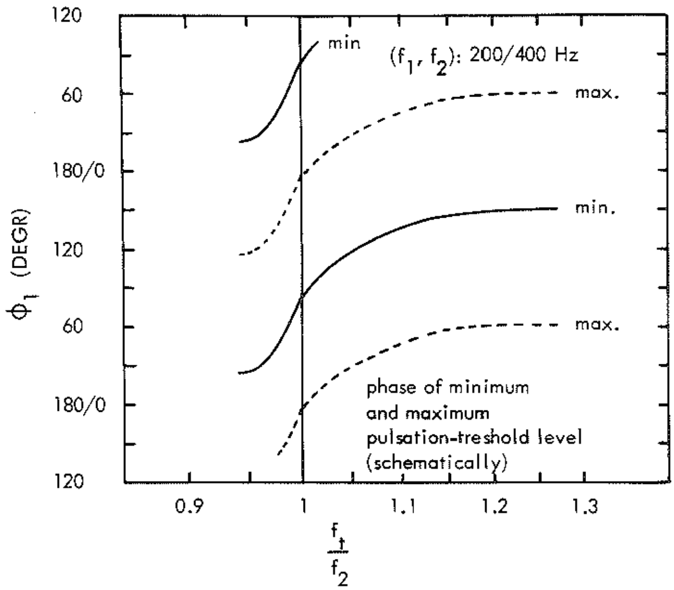


Fig. 7.4.

Schematic representation of the phase corresponding to maximum and minimum pulsation threshold as a function of the ratio of f_t and f_2 for the combination 200/400 Hz. The curves indicated with "min" are the same as in Fig. 7.2. for 200/400 Hz. The curves are repeated periodically, as the vertical scale covers two full cycles of the stimulus waveform.

The resemblance between the graph in the left-hand part of Fig. 7.5 and the pitch/phase curve for the 200/400 Hz combination (Fig. 4.1.a) is striking. If we assume that the different test-tone frequencies correspond to a spatial projection in some region of the auditory system this resemblance implies that the pitch is generated at that "place" where the pulsation threshold is maximum (as a function of phase there).

The results for the other frequency combinations show less clearly the existence of a "pitch-place relation". In these cases there is only an increase of the shift with increasing value of ϕ_1 in a restricted phase region. Nevertheless, the similarity of the pitch-phase functions for the three frequency combinations points to a similar mechanism for the production of the pitch sensation.

The following experiment was performed to find a possible "pitch-place relation" in another way. For the

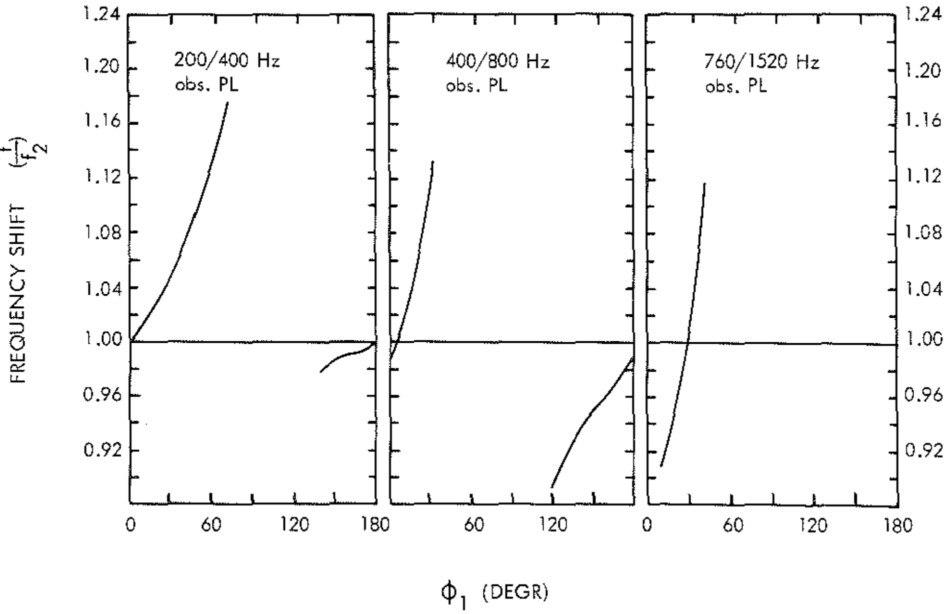


Fig. 7.5.

Schematic representation of the value of f_t relative to that of f_2 at which the pulsation threshold is maximum as a function of ϕ_1 for the frequency combinations 200/400 Hz, 400/800 Hz and 760/1520 Hz. These curves were obtained from Fig. 7.2. as explained in section 7.2.3.

frequency combinations 760/1520 Hz and 200/400 Hz we chose values of L_1 and L_2 which brought about a distinct pitch-phase effect. The pitch was matched for four values of ϕ_1: ϕ_{max} (nearly no pitch shift), $\phi_{max}+30^o$, $\phi_{max}+60^o$ and $\phi_{max}+90^o=\phi_{min}$ (maximum pitch shift). An interfering tone with frequency f_i was presented simultaneously with the complex. At the beginning of the experiment f_i was chosen so far beyond f_2 that the pitch of component II was not influenced by the interfering tone at any phase. The value of f_i was then decreased in steps of 10 Hz, and the pitch was matched at each step. The results are shown in Fig. 7.6.a and b. It may be seen that as f_i decreases the highest pitches are eliminated or influenced first. For

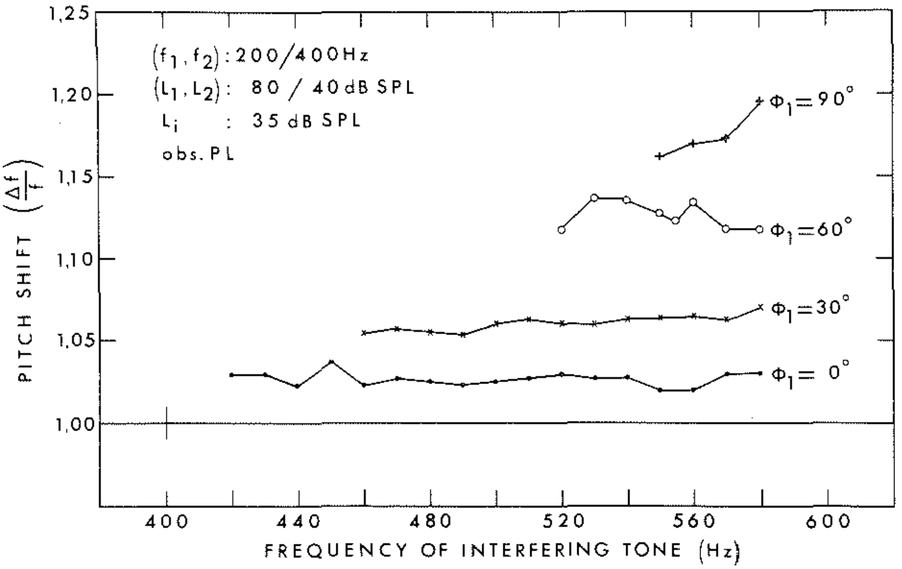


Fig. 7.6.a.

Pitch of component II of the octave complex 200/400 Hz as a function of the frequency of an interfering tone f_i. The horizontal lines correspond to the absence of pitch shift; the small vertical lines at the left-hand side of the graph correspond to the condition $f_i=f_2$.

example, in Fig. 7.6.a. for f_i=540 Hz, pitch matchings can no longer be made for $\phi_1=90^\circ$ whereas they can for $\phi_1=60^\circ$. For f_i=500 Hz, the pitch corresponding to a shift of 1.13 is also eliminated. In Fig. 7.6.b., although the effect is less distinct, for f_i=1700 Hz the highest pitch is lowered whereas the other ones are not. For f_i=1640 Hz the pitch which corresponds to a shift of 1.05 ($\phi_1=60^\circ$) is also lowered, and so on. These facts support the assumption that a "pitch-place relation" exists for both higher-and lower-frequency combinations.

However, a striking difference may be noted between Fig. 7.6.a. and b. For the 200/400 Hz combination at $\phi_1=60^\circ$ and 90°, pitch matchings cannot be made below a certain value of

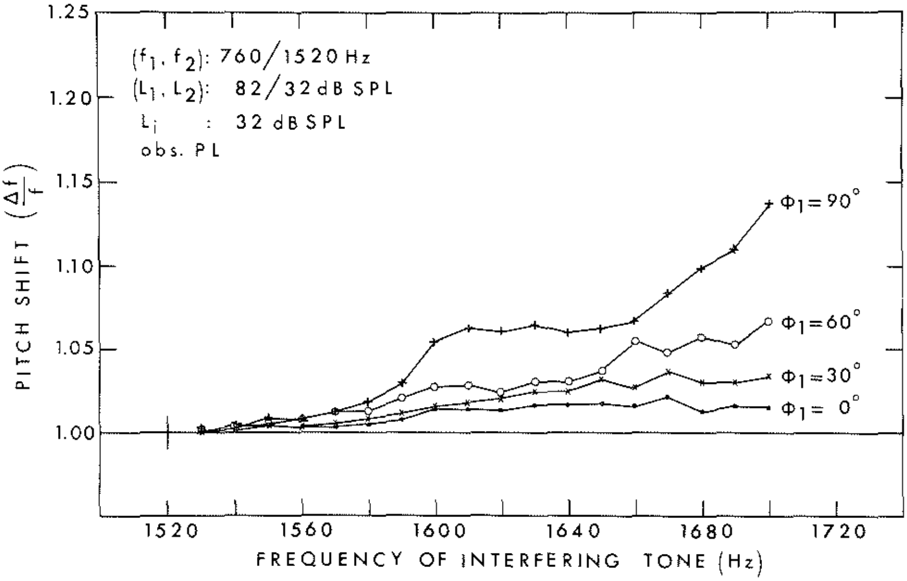


Fig. 7.6.b.
The same dependence as in Fig. 7.6.a. for the octave complex 760/1520 Hz.

f_i because component II is no longer heard separately. For the 760/1520 Hz combination pitch matching remains possible, even when f_i decreases to 1530 Hz. This discrepancy may be related to the fact that the interaction region is much wider for the 200/400 Hz combination than for the 760/1520 Hz combination (Fig. 7.2.) The interaction region for the 200/400 Hz combination overlaps the area from which the pitch is thought to originate. For the 760/1520 Hz combination the phase effects in the pulsation threshold are restricted to a narrower band around f_2 whereas the region important for the pitch generation lies either outside, or at the upper limit of this band. Assuming that the loudness effects originate from the region in which the pulsation threshold is phase-dependent, one might say that the pitch and loudness effects for the 760/1520 Hz combination are more independent of each other than they are for the 200/400 Hz combination. Affecting the pitch with an interfering tone in the latter case would mean that the loudness is affected too. Affecting the pitch for the 760/1520 Hz combination would mean that the loudness remains unaffected, in any case for not too low values of f_i.

In fact, it would be better to perform this experiment with the interfering tone presented before the stimulus, because the simultaneous presentation may induce interactions between the interfering tone and the octave complex. It was, however, not possible to obtain a sufficiently prolonged and frequency-selective "fatigue" at which pitch matching was still possible.

Summarizing, we may state that the facts presented above may provide indirect evidence that the pitch of component II originates in an area, e.g. in the organ of Corti, which is maximally sensitive to a pure tone with a frequency corresponding to the pitch. The explanation of the pitch effects will be discussed in more detail in Chapter 9, in relation to the movement of the basilar membrane.

7.3. Phase of minimum loudness as a function of frequency combination

7.3.1. *Experiments*

Fig. 4.1. and Fig. 7.1. show that the value of ϕ_1 corresponding to minimum loudness of component II depends on the stimulus frequencies. This value of ϕ_{min} was measured as a function of f_2. The results will be discussed after the presentation of the phase characteristics of the outer- and middle-ear system.

7.3.2 *Phase characteristics of the outer and middle ear*

The transfer of the stimulus from the electric signal applied to the headphone the movement of the basilar membrane involves three stages: (a) the coupling of the headphone to the concha, (b) the external auditory meatus and (c) the eardrum together with the middle-ear system. These stages are usually considered separately in the literature. The acoustics of "circumaural"* headphones has been studied by Shaw and Thiessen (1962), Shaw (1974) stated that the description of the coupling for a circumaural headphone is attended with difficulties and cannot easily be simulated with the aid of an artificial ear. Charan et al. (1965) studied the limitations of the applicability of three types of couplers for circumaural headphones. The amplitude characteristics they obtained with these couplers were not reliable over the whole audio-frequency range. Nor can the Bruel and Kjaer artificial ear, type 4153, which is available in our laboratory, be used to calibrate a circumaural headphone (Bruel et al., 1961). We may conclude that it is not possible to determine the phase characteristic of the coupling between the headphone and the concha with an artificial ear.

The sound-pressure distribution in the external auditory meatus due to a free sound field was investigated by Wiener and Ross (1946), among others. The phase characteristic of this part of the system has, however, never been measured.

The eardrum and middle ear are the best known parts of the auditory system. Their phase characteristic may be estimated via measurement of the acoustic impedance in the plane of the eardrum (see e.g. Zwislocki, 1957 and Møller, 1960).

We used the following procedure to determine the phase characteristic of the whole outer-ear system and to overcome the difficulties mentioned above. We made a plaster cast of the side of the observer's head, with the auricle and the concha. A hole was drilled in the dummy at the side of the

*We used a Grason Stadler TDH 39 circumaural headphone in our experiments.

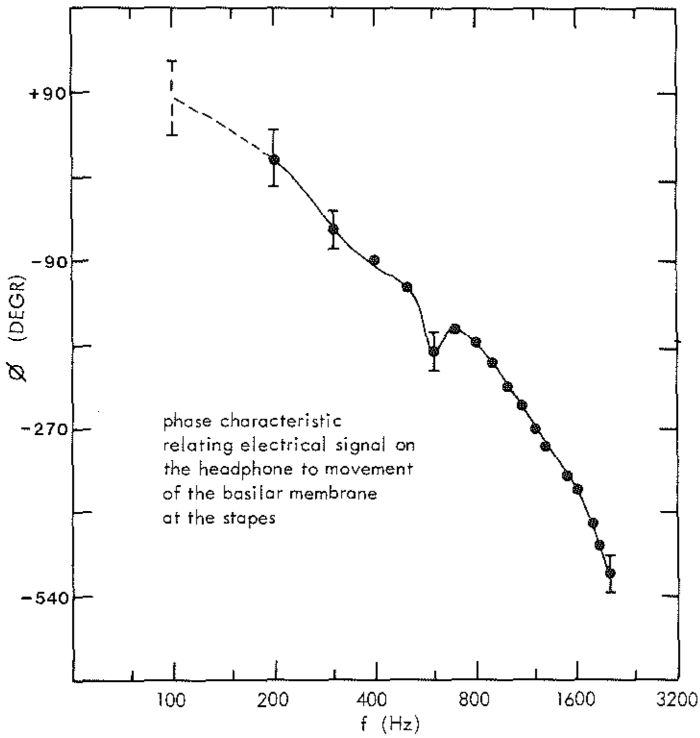
concha, perpendicular to the plane of the auricle. A brass tube of length 3 cm and diameter 0.7 cm was inserted in this hole. The tube was connected with a Bruel and Kjaer microphone at the end opposite the auricle. The headphone was placed around the auricle and pressed against the dummy by putting a 500g weight on it. Then the phase characteristic of the system between the electric signal applied to the headphone and the sound pressure in the plane of the microphone was estimated for frequencies between 100 and 2000 Hz. The choice of the upper limiting value will be discussed below. To provide some check on whether the coupling of the headphone with the cast was representative for the coupling with a real ear, a tube was stuck through the headphone cushion and connected with a probe microphone. The phase of the sound-pressure signal was measured with this probe microphone both on the cast and on the real ear. Differences between the results for the cast and for the real ear were found only for frequencies below 1000 Hz. This result is not surprising, as Charan et al. (1965) found that the acoustic signal in the concha is most sensitive to the positioning of the headphone in this frequency region. The measured phase characteristic was corrected for these differences.

Having estimated the phase characteristic for the first two parts of the system, we still need to determine the phase characteristic of the middle-ear system. For this purpose we used the phase characteristic of the middle-ear model of Zwislocki-Flanagan (Flanagan, 1962), which is based on the results of acoustic impedance measurements. We restricted ourselves to frequencies below 2000 Hz to be sure that the eardrum acts as a rigid piston, as only under these conditions is the acoustic impedance a measure of the mechanical impedance of the middle ear (Møller, 1963).

The phase characteristic obtained with the dummy together with that of the middle ear model of Zwislocki-Flanagan is assumed to relate the electric signal applied to the headphone to the movement of the stapes. To relate the latter to the movement of the basilar membrane, we must bear in mind that the motion basilar membrane has a phase lead of 90° with respect to the displacement of the stapes (Zwislocki, 1965; Rhode, 1971). The overall phase-transfer function relating the phase of the electric signal applied to the headphone to that of the movement of the basilar membrane at the stapes is given in Fig. 7.7. Displacement of the basilar membrane towards the scale vestibuli is taken as positive here.

Fig. 7.7.

Phase-transfer function between the electric signal applied to the headphone and the movement of the basilar membrane at the stapes. The displacement of the membrane towards the scale vestibuli is taken as positive.



7.3.3. *Discussion*

In section 7.3.1. we described the measurement of the phase at which the loudness of component II is minimum (ϕ_{min})

as a function of f_2 ($f_2=2f_1$). The resulting dependence of ϕ_{min} on f_2 (not shown) was transformed into a dependence of ϕ_2 on f_2 with the aid of the phase-transfer function of Fig. 7.7. The result of the transformation is shown in Fig. 7.8. The values of ϕ_2 represent the phases in the waveform of the movement of the basilar membrane at the stapes for which the loudness of component II is minimum. Note that ϕ_2 is used to describe the signals now, to facilitate comparison with data in the literature where it is more usual to describe two-component signals in terms of the phase of the higher-frequency component.

The vertical bars in Fig. 7.8. indicate the scatter of the experimental points. It may be seen that the minimum loudness of component II does not correspond to one phase relation in the waveform at the basal end of the basilar membrane for all frequency combinations. The implications

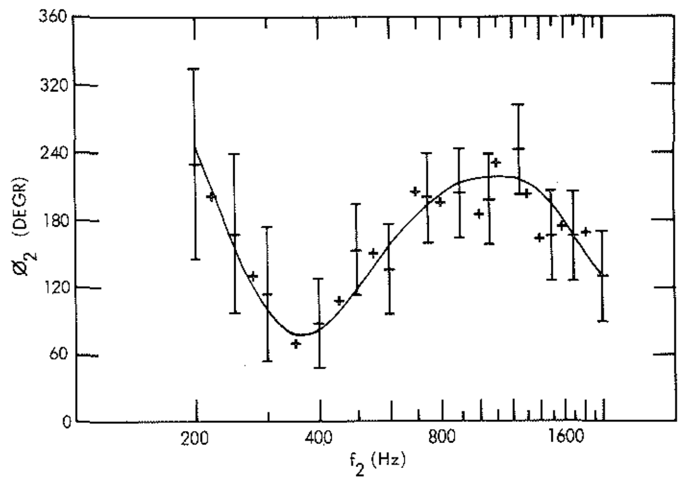


Fig. 7.8.

Value of ϕ_2 in the waveform

$$f(t)=A_1\sin(2\pi f_1 t) + A_2\sin(2\pi 2f_1 t+\phi_2)$$

of the movement of the basilar membrane at the stapes that correspond to minimum loudness of component II, as a function of $f_2=2f_1$. A positive value if f(t) corresponds to a displacement of the membrane towards the scala vestibuli.

of this result will be discussed more thoroughly in Chapter 9 together with the results of the first part of this chapter.

7.4. Conclusions

1) The phase condition under which the pulsation threshold is minimum or maximum is a function of the test-tone frequency for a given frequency combination. The corresponding value of ϕ_1 increases with increasing f_t.
2) The range over which the phase varies as a function of f_t decreases from 120^o for the combination 200/400 Hz to 30^o for 760/1520 Hz.
3) The width of the interaction region for a given octave complex, i.e. the range of f_t over which the phase is found to influence the pulsation threshold varies as a function of f_2 in the same way as the critical band.
4) The pitch effects are subject to a "pitch-place relation". This implies that the pitch perceived originates in a region that is maximally sensitive to the frequency corresponding to the pitch.
5) For the 200/400 Hz frequency combination the pitch corresponds to the value of f_t at which the pulsation threshold has the maximum value that, as a function of the phase, can be found for that f_t.
6) The phase at which the loudness of component II is minimum does not correspond to one and the same phase in the waveform of the movement of the basilar membrane at the stapes for the different frequency combinations.

PART IV

IMPLICATIONS OF THE EXPERIMENTAL RESULTS

CHAPTER 8

IMPLICATIONS OF THE RESULTS FOR THE AURAL-HARMONICS HYPOTHESIS

8.1. Introduction

Let us now return to the questions formulated in section 2.5. The first question concerns the precise nature of the sensations involved in the perception of the higher-frequency component of two-tone octave complexes. Our attempts to answer this question are described in Chapters 4 and 5.

In Chapter 4 data were presented on effects of phase on the pitch and the loudness of component II. We saw that these effects are highly interdependent, in such way that the jump observed in the pitch coincides with the minimum in the corresponding loudness. It was argued that the pitch effects cannot be caused by the variation in loudness. Furthermore, we saw in Chapter 5 that the variations in the pitch, the pulsation threshold and the masked threshold for on-ratio octave complexes are related to those for off-ratio complexes. This supports the idea that two-tone stimuli give rise to a continuum of interaction effects, which are phase-dependent for octave complexes.

The facts outlined above also provide an answer to the second question posed in section 2.5., as to whether component II can be considered as an independent probe tone which can be used for detection of possible aural harmonics. On the basis of our experimental results, the answer is "no", even for values of L_2 below the "best beats" range (see Fig. 2.1).

8.2. To what extent can the phase effects be understood in terms of vector summation of an aural harmonic and component II?

The rest of this chapter is devoted to answering the third question formulated in section 2.5: whether the phase effects can be understood in terms of vector summation of an aural harmonic and component II. First, we will find out whether there is general evidence for aural harmonics and what is the contribution of two-tone nonlinearity in this connection.

8.2.1. *Evidence for aural harmonics. Possible site of their generation*

As far as we are aware, the first study on the linearity of the eardrum and ossicles under physiological conditions was performed on the cat by Guinan and Peake (1967) with light microscopy in stroboscopic illumination. Unfortunately, their method was not sufficiently sensitive to show possible distortion below - 20 dB. Further indications that the response of the eardrum and middle ear is linear can be found in the data of Rhode (1971) who applied the Mössbauer technique in the squirrel monkey, and Wilson and Johnstone (1975) who applied a capacitive probe technique in the guinea pig. Moller (1974) concludes that, if the conditions found in the cat are valid for men too, the middle ear can be regarded as a linear element which does not produce any distortion components of importance for the perception of ordinary sounds.

Several authors have investigated whether the basilar membrane has a linear or a non-linear response. Rhode (1971), Rhode and Robles (1974) and Robles et al. (1976) found nonlinearities near the vibration maximum for levels between 70 and 110 dB SPL at a frequency of 7 kHz. Such nonlinearities were not found by von Békésy (1949), Johnstone and Boyle (1967), and Wilson and Johnstone (1975). There is a clear discrepancy between the results of Rhode and co-workers and those of Wilson and Johnstone. So far, it is not known whether this is due to a difference in the choice of the experimental animal or to a difference in experimental technique.

Models which can describe cochlear nonlinearities were developed by Tonndorf (1957, 1958a, 1958b, 1959), Kim et al. (1974), Hall (1974) and Schroeder (1975). Tonndorf using a scale model of the cochlea, made an extensive study of fluid movement in the scalae and its relation to membrane displacement. The displacement pattern of the basilar membrane shows peak clipping if stimulated by a "pure tone" at an equivalent level above 80 dB SPL. This is caused by "Bekesy eddies". The harmonics generated travel back to the places on the basilar membrane corresponding to their own particular frequencies. Tonndorf (1958b) demonstrated such travelling waves in the cochlear microphonics of guinea-pig cochleas at levels of 90 dB SPL and above. However, Rhode (1971) found no peak clipping in the squirrel monkey. Furthermore, Wilson and Johnstone (1975) found the displacement of the guinea-pig basilar membrane to be linear. Finally, Dallos and Sweetman (1969), measuring cochlear microphonics in the guinea-pig, found that harmonics at low levels (60 dB SPL) do not give rise to travelling waves but are restricted to the site of the fundamental. Under these level conditions, distortion is attributed to the mechano-electric conversion process in the hair cells. At higher levels the results (Dallos, 1973) do not appear to show any clear tendency to favour either a localized process or travelling waves. The over-all experimental evidence, therefore, does not seem to favour Tonndorf's findings.

Kim et al. (1974), Hall (1974) and Schroeder (1975) postulated a nonlinear relation between membrane velocity and pressure. Their model describes the nonlinearity found by Rhode (1971) and the behaviour of the cubic difference tone. Even-order distortion products can be obtained with this model by introducing an asymmetrical relationship between current and resistence. Robles et al. (1976) did not describe in detail the nonlinearity they found, but only stated that the model gives good agreement with their results. The question of whether mechanical nonlinearity can give rise to even-order distortion products or whether aural harmonics are generated by another kind of mechanical nonlinearity is thus still open.

In spite of the fact that aural harmonics can be generated so easily in a model, single-cell recordings in the auditory nerve and the cochlear nucleus have never revealed any indication of the existence of aural harmonics so far, either in the form of peaks in patterns such as "tuning curves" (Kiang et al., 1965) or as a contribution to period histograms (Rose et al., 1974).

Summarizing, we may safely conclude that aural harmonics, if present, are not generated in the eardrum or the middle ear. Generation in the mechanical part of the cochlea cannot be excluded. There must, however, remain doubt about such generation because of Wilson and Johnstone's results (1975) and because there is no evidence for their presence from single-cell recordings. It may be that aural harmonics, once generated, are suppressed due to "two-tone suppression". We will discuss this briefly in the next section.

8.2.2. *Aural harmonics in relation to two-tone interaction*

The aural harmonics hypothesis involves the description of the phase effects in terms of vector summation of component

II and an assumed aural harmonic. Direct application of the model would imply that the loudness of component II when $\phi_1=\phi_{max}$ is higher than that of component II in the absence of component I. Our loudness data confirm this at lower values of L_2; but at higher values of L_2 we generally found that the loudness at $\phi_1=\phi_{max}$ is lower than the original value (Fig. 6.4). Similar conclusions can be drawn from the results of the pulsation-threshold experiments. The curves for L_2= 25 dB SPL in Fig. 6.2. agree with the picture of vector summation, while those for L_2= 45 dB SPL do not.

We concluded in Chapter 6 that the pulsation threshold and loudness results are subject to "two-tone suppression" and "two-tone enhancement" effects. If we assume that aural harmonics are involved in phase effects, we may, therefore, conclude that they are also influenced by two-tone suppression and enhancement. This view, which as far as we know has never been expressed in the discussion about aural harmonics in the literature before, provides a second argument against the supposition that component II can be regarded as an independent probe tone.

In Chapter 3 we gave arguments why, in our opinion, the pulsation-threshold method is the best technique available at the moment for study of two-tone interaction since the results obtained with it resemble single-cell recordings from the auditory nerve and cochlear nucleus. In the present discussion we shall examine whether peaks are present in the pulsation threshold patterns of pure tones at $f_t=2f_1$ (see section 8.2.4). Before doing so, however, we will review what is known about the localization of the two-tone interaction.

8.2.3. *The site of two-tone interaction*

The place along the basilar membrane or in the organ of Corti where two-tone interaction might take place is still unknown. Most investigators assume the source of two-tone nonlinearity to be localized after the process of mechanical filtering. Furman and Frischkopf (1964) suggested that it might be due to interaction of receptor-neuron elements. Pfeiffer (1970) assumes the site of two-tone nonlinearity after the mechanical filter. Duifhuis (1976) situates it in the transfer of the movement of the membrane into that of the hairs. On the other hand, Schroeder (1975) assumes two-tone nonlinearity to originate from a nonlinear loss mechanism in the movement of the basilar membrane. There are a number of reasons for disagreeing with Schroeder. In the first place, we concluded in section 8.2.1. that evidence is available that the response of the basilar membrane is linear. Secondly, our own results show that suppression is already present at a relatively low level (L_2= 45 dB SPL; see Fig. 6.2). A third objection is that the pulsation threshold first decreases and then increases again with increasing value of L_1. If we take these objections into account, it seems justified to localize the two-tone nonlinearity after the mechanical filtering.

8.2.4. *Which comes first, two-tone interaction or generation of aural harmonics (if present)?*

We have seen in the above sections that the opinions about the sites of a possible mechanical nonlinearity and of two-tone nonlinearity are far from unanimous. Nevertheless, we may select two possible sequences of these two processes as being most likely. One is generation of aural harmonics in the mechanical part of the cochlea followed by the two-tone nonlinearity. The other is that aural harmonics (if present) are generated at the same stage as two-tone nonlinearity, after the mechanical filtering.

We will discuss these two possibilities in turn, without considering whether aural harmonics are actually present, but only whether the phase effects can be described in terms of aural harmonics, if they exist.

Generation of aural harmonics in the mechanical part of the cochlea implies that component I and the vector sum of the aural harmonic and component II together form a new two-tone stimulus which is subject to suppression and enhancement. The data presented above enable us to check this statement.

We repeat our assumptions. First, the assumed aural harmonics are considered to be due to mechanical distortion in the ear. The second assumption is that the two-tone nonlinearity is introduced after the mechanical distortion stage. Our third assumption is that we consider the pulsation threshold as representing the output of a system consisting of the two above-mentioned stages in the order indicated.

Our aim is to estimate the levels of these assumed aural harmonics with the aid of Fig. 6.2, taking the nonlinear phenomena into account. On the basis of these results we can decide whether one may expect a peak in the pulsation patterns of the corresponding lower-frequency components alone.

Let us consider Fig. 6.2. It is reasonable to suppose that the values of L_1 and L_2 at which the difference in pulsation threshold between ϕ_{max} and ϕ_{min} is maximum correspond to conditions under which the mechanical levels of the aural harmonic and the higher-stimulus component are equal. For instance, complete cancellation takes place for L_1= 70 dB SPL and L_2= 25 dB SPL. The mechanical level of the aural harmonic must therefore correspond to 25 dB SPL. These mechanical levels are given in the second colum of Table 8.I. The corresponding values of L_1 are given in the first column. We note that the level of the assumed aural harmonic rises about three times as fast as the value of L_1.

L_1	equivalent level of mechanical aural harmonic	contribution of component I without aural harmonic	equivalent level of aural harmonic + component II at $\phi = 10^{o}$	pulsation threshold at $\phi = 10^{o}$	suppression	output level of aural harmonic
70 dB SPL	25 dB SPL	≤ 22 dB SPL	31 dB SPL	32 dB SPL	-1 dB	26 dB SPL
74 dB SPL	35 dB SPL	≤ 22 dB SPL	41 dB SPL	35 dB SPL	+6 dB	33 dB SPL
77 dB SPL	45 dB SPL	32 dB SPL	51 dB SPL	39 dB SPL	+12 dB	-*

Table 8.I.

*could not be determined for reasons given in the text.

Calculation of the output levels of the hypothetical aural harmonic needed to describe the data of Fig. 6.2. The calculation is explained in section 8.2.4. The different columns contain the following data:

1) The values of L_1 for which the calculation is performed.
2) The equivalent mechanical levels of the hypothetical aural harmonic.
3) The contribution of component I after the effect of the aural harmonic has been cancelled out.
4) The equivalent mechanical levels of the aural harmonic added at the correct phase to component II ($\phi_1 = \phi_{max} = 10^{o}$.
5) The pulsation threshold at $\phi_1 = 10^{o}$
6) The suppression (the difference between the values in columns 4 and 5). A negative suppression indicates "enhancement".
7) The output levels of the hypothetical aural harmonic.

This does not fit in with the idea of quadratic distortion.

Fig. 6.2. shows that the cancellation of the assumed aural harmonic by component II is not complete. The residual masking under the cancellation conditions (e.g. for L_1=77 dB SPL, L_2=45 dB SPL and ϕ_1=100^{o}) may be interpreted as an output due to the flat slope of the stimulation pattern of component I. This contribution for the three values of L_1 is given in the third column of Table 8.I.

Reinforcement of the assumed aural harmonic and component II (ϕ_1=10^{o}) for the value of L_1 in the first column of Table 8.I. leads to an increase of 6 dB in the mechanical level. These values are given in the fourth column. The corresponding pulsation thresholds are given in the fifth column. The difference between the fourth and fifth columns yields the degree of suppression, which is given in the sixth column.

A negative value here indicates enhancement.

We must subtract 6 dB from the values given in the fifth column to obtain the output levels of the aural harmonic during the presentation of the lower-frequency component alone. However, this reduction is generally coupled with a fall-off in suppression. For L_1=70 dB SPL there is no suppression, so the output level of the aural harmonic corresponds to 32-6=26 dB SPL. This value is given in the seventh column of Table 8.I. For higher values of L_1 a decrease in L_2 by 10 dB corresponds to a decrease in suppression by 6 dB. Therefore, for L_1=74 dB SPL the resulting decrease of the output is 2 dB ($6-\frac{6}{10}\cdot 6$ dB) yielding 35-2=33 dB SPL. This reasoning does not hold for L_1=77 dB SPL, as the output level given in column 4 (39 dB SPL) contains the assumed aural harmonic together with a considerable contribution from the lower-frequency component itself (33 dB SPL; column 3). Therefore, subtracting 2 dB from the value of 39 dB SPL in column 5 would yield too high a value for the aural harmonic.

The results given in column 7 show that the difference between the output due to the assumed aural harmonic and that due to component I without the aural harmonic (column 3) is maximum for L_1=74 dB SPL (33 as compared with $\leq$22 dB SPL). This means that we would expect a peak in the pulsation pattern of a 400 Hz tone of L_1=74 dB SPL in the test-tone frequency region around f_2. This pattern was determined by the same observer as Fig. 6.2., and is shown in Fig. 8.1. (full line). There is no evidence of a peak. The absence of peaks at $f_t=2f_1$ was noticed earlier by Munson and Gardner (1950) who measured masking patterns of pure tones with forward masking.

It must thus be concluded that our experimental data offer no support for the view that the phase effects are caused by mechanical aural harmonics. This conclusion is

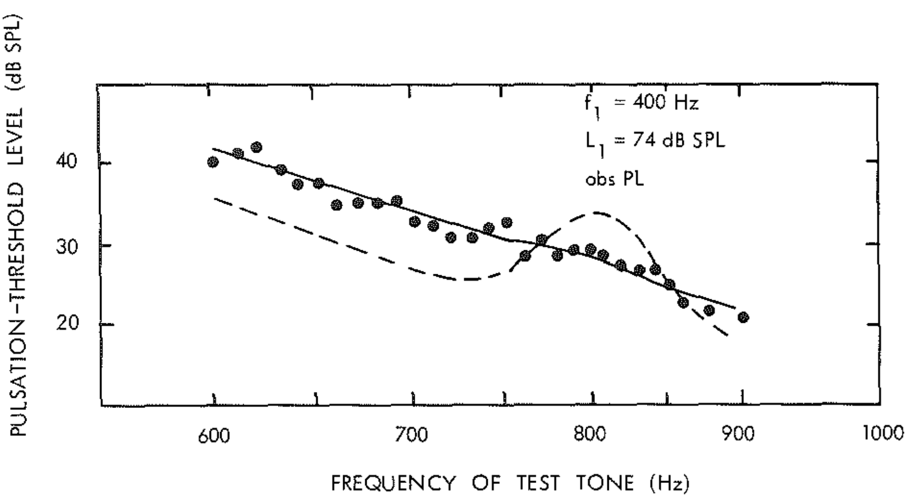


Fig. 8.1.

The full line represents the pulsation threshold of a pure-tone stimulus in the test-tone frequency region about one octave above the frequency of the masker. F_1=400 Hz, L_1=74 dB SPL. The broken curve indicates the result that would be expected if the effects of phase on the pulsation threshold level were caused by mechanical aural harmonics (see section 8.2.4).

supported by the fact that no beats can be heard when the octave complexes are presented dichotically (see Chapter 4). Consequently we must conclude that our phase effects cannot be explained on the assumption of vector summation in combination with two-tone interaction.

The remaining possibilities are either that a second harmonic is generated at the same stage as that at which the two-tone nonlinearity occurs, or that the two-tone nonlinearity brings about phase effects in the pulsation threshold and the loudness without a second harmonic playing a rôle. This implies that the phase effects must be generated at a stage after mechanical filtering, e.g. in the hair-cell/neuron transducer. In any case, a possible explanation of these effects would be incomplete if it did not explain the pitch effects too.

Clack et al. (1972) applied the vector-summation model to the effects of phase on the masked threshold of component II and estimated the levels of the assumed aural harmonic as a function of L_1. In the masking method used by these authors, the masker (component I) and the test tone (component II) are presented simultaneously. This has the disadvantage that the two-tone nonlinearity is overlooked (see also Chapter 3). The arguments given above indicate that the levels calculated by Clack et al. do not reflect mechanical aural harmonics. If the phase effects originate from two-tone nonlinearity, the results of Clack et al. need to be reinterpreted.

8.2.5. *Discussion of findings which seem to be in favour of vector summation*

In the preceeding sections we concluded that the phase effects cannot be described in terms of vector summation of component II and an aural harmonic. This fact together with the absence of beats when the octave complexes are presented dichotically indicates that aural harmonics, if present, do not behave like externally presented tones. However, there are some phenomena which at least at first sight, do seem to be in agreement with the concept of vector summation. We review them below in order to find out whether or not they really are in disagreement with the above conclusion.

First, the form of the curves for L_2=25 dB SPL in Fig. 6.2. is very much like what we would expect in the assumption of vector summation, although no corresponding peak is present in the pattern for component I alone. Further, very reasonable levels of "aural harmonics" can be estimated from the results of tone-on-tone masking experiments, even though these harmonics are not in fact present.

Another point of agreement is found in an experiment performed by Schouten (1938). He minimized the strength of the beats in the frequency combination 200/406 Hz by adding a 400 Hz tone with an appropriate phase. The experiment was repeated and the result confirmed by Schoonderbeek (1976) for six observers and different frequency combinations. Both these authors interpreted their results as indicating the existence of aural harmonics; but the existence of such an effect can also be explained in the light of our experimental results by assuming that two types of beats are present: that of the combination (f_1, $f_2+\Delta f$) and that of the combination (f_2, $f_2+\Delta f$) which oppose one another.

Schouten (1938) performed another experiment, the result of which seems to support not only the idea of vector summation but even the audibility of "aural harmonics". Schoonderbeek (1976) also repeated this experiment with different observers and frequency combinations. The experimental procedure is as follows. A continuous component I and a pulsating component II (200 ms on and 200 ms off) are presented together. A phase and a value of L_2 can be found at which the observer no longer perceives the addition of component II but rather the removal of something. Continued listening results in a shift of the attention so that one has the impression that in the intervals in which only component I is present a tone with a pitch roughly corresponding to a frequency f_2 is added. If component II is switched off, this higher-frequency tone remains audible as a continuous tone for a few moments.

The effect can be understood quite easily in the light of our pulsation-threshold data. Fig. 6.2. shows that, under specific level and phase conditions, the pulsation threshold for the complex at $f_t=f_2$ is lower than that for component I alone. Under these conditions a gap was found in the pulsation pattern of the complex around $f_t=f_2$, whereas such a gap was absent in the pattern of component I alone. Alternate presentation of these two patterns focusses the attention on the area which varies: the frequency region around f_2. The pattern without the gap (component I) can then be interpreted as the addition of a sensation corresponding to the frequency f_2.

On the other hand, some of our observers reported that they could hear a second harmonic in a continuous pure tone directly, without the addition of a component II as in Schouten's experiment. The "harmonic" was perceived as being slightly too high in pitch with respect to the frequency f_2.

It is possible that such sensations exist without observers being clearly aware of them. Our conclusion formulated in the previous section only implies that, if present, they do not originate from mechanical distortion but from a stage after the mechanical filtering. In any case, our findings are not necessarily in contradiction with Plomp's (1967) conclusion that it is possible to hear aural harmonics, but by no means everyone possesses this ability. (see section 2.4.2).

In conclusion, we may state that the experimental results we have reviewed, which were regarded as supporting the concept of vector summation, can also be interpreted in other ways.

8.2.6. *The simultaneously-masked threshold of component II as a function of L_1*

Plotting the simultaneously-masked threshold of component II (component I as masker) as a function of L_1, gives a curve with a slope 2 for some frequency combinations. This slope has been often considered as an indication of quadratic distortion (see Chapter 2). We have two arguments against this reasoning. First, a slope of two is only found with simultaneous masking when two-tone interaction is overlooked. Secondly, we found that the slope varies with the frequency of the masker (Table 5.II). Phase effects are also found when the slope differs from 2 (see also Nelson and Bilger, 1974).

8.3. Conclusions

1) The effects of the phase on loudness and pulsation threshold cannot be explained in terms of mechanical aural harmonics, for the following reasons:

a) the aural harmonics needed to describe the phase effects when two-tone suppression is taken into account are absent in the pulsation patterns of the lower-frequency component alone.

b) no beats can be heard when the octave complex is presented dichotically (one component to each ear).

2) The phase effects, and sensations of second harmonics which some observers seem to be able to hear, probably originate in a stage after mechanical filtering, e.g. in the hair-cell/neuron transducer.

CHAPTER 9

IMPLICATIONS OF THE RESULTS FOR THE WAVEFORM HYPOTHESIS

9.1. Relation between the dependence of the phase for minimum (maximum) pulsation threshold on f_t and the waveform of the basilar-membrane motion

The last question formulated in section 2.5 concerned the possible relation between the observed phase effects and variations in the stimulus waveform as a function of phase. We shall now consider whether such a relation does exist, with reference to the dependence of the phase for minimum or maximum pulsation threshold on f_t (Fig. 7.2) and the dependence of the phase for minimum loudness on the frequency combination (Fig. 7.8).

Let us first look at the curve for the 200/400 Hz frequency combination in Fig. 7.2. A variation of this form can be expected if we assume that the minimum and maximum pulsation thresholds each correspond to a waveform on the basilar membrane with a specific phase relation between the components. This can be demonstrated qualitatively with the aid of the phase characteristic of the motion of the basilar membrane, as determined by Rhode (1971). This characteristic shows a linear relation between phase lag and frequency for frequencies which are low compared with that frequency for which the site under investigation is maximally sensitive (7.1 kHz). This implies that the velocity of the travelling waves on the basilar membrane is independent of frequency. Thus, the phase relation in the travelling wave, due to an octave complex will not alter along the membrane on the stapes side (i.e. the side away from the peak corresponding to component II). Indeed, the curve of Fig. 7.2. reaches a plateau for values of f_t much larger than f_2. When, however, the frequency is increased till it approaches that for maximum sensitivity, the phase lag in the phase characteristic measured by Rhode increases faster than corresponding to the linear relation for lower frequencies. This implies that near the place of maximum sensitivity for component II the phase relation in the travelling wave alters in such a way that the phase of component II decreases going from stapes to apex. This in its turn implies that the value of ϕ_1 in the stimulus waveform has to be decreased in order to maintain the same waveform at all places along the basilar membrane. If the minimum and maximum pulsation thresholds each correspond to a specific waveform, the value of ϕ_1 at which the minimum or maximum occurs must thus decrease with decreasing f_t, which is indeed found to be the case.

The correspondence of both the minimum and maximum pulsation thresholds to constant phase relations in the waveform along the basilar membrane is less marked for the combinations 400/800 Hz and 760/1520 Hz than it is for the combination 200/400 Hz. Nevertheless, the same trend is found. The results for a frequency ratio 1:3 with f_2=400 Hz were quite different, however: the pulsation threshold could be minimized over a wide test-tone frequency range (the ratio between f_{tmax} and f_{tmin} being 2.3) but the phase at which this occured did not depend on f_t.

We can imagine two effects that might be responsible for these differences between the results for the different frequency combinations. First, we may assume the scanning of the varying waveform along the basilar membrane involves some kind of averaging of the different waveforms within a region. The variation of the size of the interaction region indicated in Fig. 7.3. may be taken as evidence for this. When the area within which the averaging occurs increases, the accuracy of the "scanning" will decrease. This will result in a decrease of the range over which the value of ϕ_1 at which the pulsation threshold is minimum varies as a function of f_t. On this assumption, it is plausible that the phase at which the pulsation threshold is minimum for f_1:f_2=1:3 and f_2=400 Hz does not depend on f_t, since the width of the interaction region here corresponds to a factor 2.3 as compared with a factor 1.3 for the 200/400 Hz combination.

Secondly, the part of the phase characteristic which is scanned may be different for the different frequency combinations. For the higher-frequency combinations this part would be near the beginning of the characteristic, where there is a linear or a nearly linear relation between phase lag and frequency. For lower frequency combinations it could be nearer the end of the characteristic where a clear deviation from the linear relation for the beginning of the characteristic exists. We have not been able to check the validity of the two suppositions.

The deviation from a linear relation between phase and frequency can be identified as dispersion, i.e. the effect that the velocity of the travelling wave depends on frequency. It is commonly found in the results of measurements of the motion of the basilar membrane (Rhode, 1971), and Wilson and Johnstone, 1975). In electrophysiological data, however, it is sometimes absent. Anderson et al. (1971) found a linear relation between the phase lag of the nerve discharges and the stimulating frequency in auditory nerve fibres of the squirrel monkey, even for a unit with a "best frequency" of 200 Hz. On the other hand, Pfeiffer and Molnar (1970) did find a dispersion-like relation for the discharges in cochlear fibres of the cat for units with best frequencies above 1000 Hz. Below 1000 Hz they found either a linear relation or even the opposite effect: a slower increase in the phase lag when the stimulating frequency increased up to the best frequency. Like our psychophysical data, these electrophysiological results do not reveal an unequivocal relation with the mechanical phase characteristic.

9.2. What waveform corresponds to minimum loudness?

It may be concluded from Fig. 7.8. that neither the minimum nor the maximum loudness corresponds to a constant waveform on the basilar membrane at the shapes for all frequency combinations. However, it must be realized that both $\phi_1=\phi_{min}$ and $\phi_1=\phi_{max}$ correspond to the situation for the pulsation threshold at $f_t=f_2$. This does not reflect the situation at the basal end of the membrane. Still, the phase relation continues to depend significantly on frequency even after correction for this difference. The dependence might equally well be due to the effects suggested in the previous section. If the area over which the varying waveform is averaged varies, the averaged phase may change as well. Similarly, differences in the averaged phase can appear when the interaction region covers different parts of the phase characteristic for the different frequency combinations. Anyway, we cannot indicate one particular phase which gives rise to either minimum (maximum) pulsation-threshold level or minimum (maximum) loudness.

The values of the phase of component II at the entrance of the external meatus for which the loudness of component II is minimum (ϕ_{2a}) or the masked threshold is maximum, as reported in the literature, reveal a wide variation. Nelson and Bilger (1974) reported a value of 90^{o} for all frequency combinations. Terhardt and Fastl (1971) found a value of 90° for the combination 200/400 Hz and Kuriyagawa and Kameoka (1966) a value of 82° for the combination 440/880 Hz. De Boer and Bouwmeester (1975) reported values decreasing from around 360° for the combination 400/800 Hz down to 60° at 1200/2400 Hz. This decrease is in agreement with our results above 500/1000 Hz. The variation in the values of ϕ_{2a} found in the literature may be due to the use of different headphones and cushions. Some authors mentioned that the phase shift between the electric and acoustic signal was measured by means of an artificial ear, which makes mutual comparison difficult (see Chapter 7).

Rose et al. (1974) studied the discharges in single units of the cochlear nucleus of the cat. For the frequency combination 380/760 Hz the period histogram for minimum discharge rate could best be described by the equation

$$f(t) = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi 2 f_1 t + 90^{\circ}).$$

It is not known whether this phase relation is frequency dependent and what the relation is with the phase in the acoustic signal.

9.3. Implications for the pitch effects

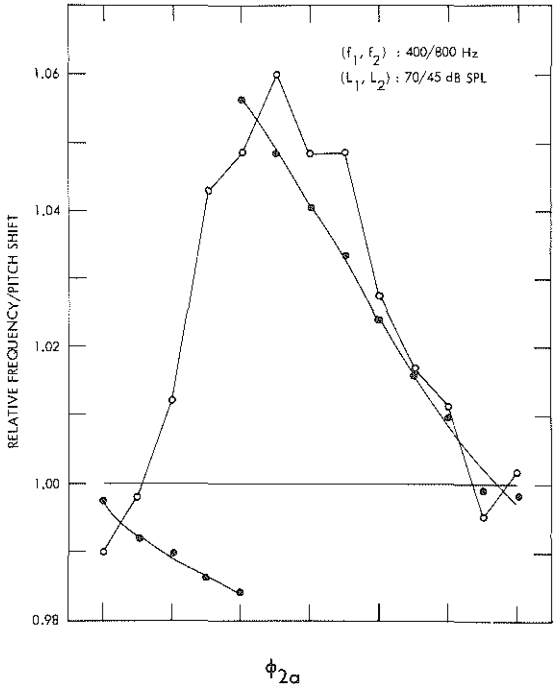
The results presented in Chapter 7 revealed that the effect of phase on the pitch of component II is caused by a shift in the place determining the pitch in some internal representation. The pitch perceived corresponds to the frequency of a pure tone for which that place is maximally sensitive. We also saw in Chapter 7 that the pitch of component II in the 200/400 Hz frequency combination corresponds to the value of f_t at which the pulsation threshold has the maximum value belonging to that test-tone frequency. Furthermore, for this combination the maximum pulsation threshold corresponds to a specific waveform. Combining these two facts and considering f_t in terms of place along the basilar membrane, we may conclude that the pitch corresponds to the place along the basilar membrane at which a specific waveform is found. The pitch is determined by the frequency of the pure tone corresponding to that place. This conclusion holds for the 200/400 Hz frequency combination.

It should be further remarked that we do not know which waveform is involved here, for the reasons explained in section 9.2. Moreover, it remains an open question whether the characteristic of this specific waveform which makes it suitable for determining the pitch is that it evokes a maximum spike rate in the connected neurons (reflected in our pulsation-threshold measurements) or that it conditions the spikes in some other way, e.g. by giving them a special time structure. This question was already involved in Chapter 6, where we failed to find a shift of a peak in the pulsation pattern as a function of ϕ_1. Patterns measured with forward masking, however, do show such a shift in a restricted phase region, as was found by Schoonderbeek (1976) for the 400/800 Hz frequency combination. The maxima in these patterns were determined by means of an averaging technique. The values of f_t corresponding to these maxima are plotted against phase in Fig. 9.1. (circles). In addition, Schoonderbeek measured the pitch of component II as a function of the phase for the same frequencies and levels. These results are also shown in Fig. 9.1. Although there is a partial similarity between the two curves, the f_t curve lacks the "jump" which is so consistently found in the pitch. This would seem to contradict the view that the pitch effects are due to a shift of a maximum in the internal representation as measured either by the pulsation-threshold method or by forward masking.

The correspondence of the pitch with the place on the basilar membrane at which a specific waveform is found may indicate why two different pitches can sometimes be heard

Fig. 9.1.

Comparison of the pitch of component II with the value of f_t at which, averaged over some test-tone frequency range a maximum is present in the forward-masking pattern of the stimulus as a function of ϕ_{2a}. Frequency combination 400/800 Hz. Level combination 70/45 db SPL. The two experiments were carried out by the same observer (from Schoonderbeek, 1976).
●● Pitch as a function of phase.
oo F_t as a function of phase.



simultaneously. If there is a large phase variation over the area corresponding to component II, the specific waveform can be present at different places. The fact that the pitch originates from the corresponding "places" agrees with current ideas on pitch perception, i.e. with the "internal spectrum" concept (see e.g. Houtsma and Goldstein, 1972 and Bilsen and Goldstein, 1974).

Our experimental results are not in agreement with the traditional time theory of pitch, in spite of the sawtooth-like variation of the reciprocal of the time between two successive peaks in the complex waveform as a function of phase. In the complex waveforms of the off-ratio stimuli mentioned in Chapter 5 the interpeak distances become larger and larger with increasing L_1. According to the time theory this would imply a decreasing pitch. However, the pitch is actually found to increase (Fig. 5.1). The relation found between the pitch effects in the on-ratio and off-ratio situations allows us to extend this conclusion to octave complexes and, generally (on the basis of the monotonic course of the curves in Fig. 5.3) to all two-tone stimuli.

Another explanation of the pitch shifts in terms of a varying "place" has been given by Egan and Meyer (1950) and, more recently, by Terhardt (1972). Terhardt uses the concept of "excitation pattern" ("Erregungsverlauf"). This should not be confused with the above-mentioned "internal spectrum", or "central excitation pattern". The "excitation pattern" as discussed by Terhardt is obtained from the simultaneous-masking pattern of a stimulus by plotting the excitation level as a function of the "tonalness" ("bark"). The explanation starts from the fact that the pitch effects occur only under conditions of partial masking. The masking of - in our case - component II by component I implies that a certain part of the excitation pattern belonging to component II on the low-frequency side of the pattern is removed by the presence of component I. The size of this excitation pattern diminishes more and more as L_1 increases, so that the maximum of the remaining excitation pattern shifts to higher frequencies. Assuming a place-pitch relation, this results in an upward pitch shift.

This explanation has the advantage that use is made of the place-pitch relation. Moreover, the fact that partial masking is a necessary condition for the pitch shifts (Chapter 5) is taken into account. However, the rôle of the excitation patterns in the explanation should be considered with reserve for three reasons. Firstly, partial masking in octave complexes occurs without any pitch shift for $\phi_1=\phi_{max}$ (Fig. 4.4). Secondly, seperate pitches can be heard simultaneously. Thirdly, the fact that there is a maximum in the pitch shift for frequency ratios 1:2 and 1:3 is not easy to understand in the light of a shift of the excitation pattern.

9.4. Conclusions

1) For low-frequency combinations such as 200/400 Hz the maximum and the minimum in the pulsation threshold as a function of the phase each correspond to a specific waveform along the basilar membrane. This conclusion is based on the qualitative agreement between the variation of the phase at which the pulsation threshold is minimum as a function of f_t and the variation of the phase relation between the two components of the mechanical wave on the basilar membrane, due to dispersion.

2) For low-frequency combinations the pitch originates from the area in which the waveform that gives rise to the maximum pulsation threshold is situated. The pitch corresponds to the frequency of the pure tone for which this area is maximally sensitive.

3) The relation mentioned in conclusion 1 is less marked for the frequency combination 400/800 Hz and is nearly absent for the combination 760/1520 Hz. However, the trend is the same.

4) Neither the minimum loudness nor the maximum pulsation threshold corresponds to a specific phase relation in the waveform of the motion of the basilar membrane at the stapes for all frequency combinations.

PART V

DYNAMIC ASPECTS AND MODEL

CHAPTER 10

PULSATION THRESHOLD IN RELATION TO FORWARD MASKING. EVIDENCE FOR THE PRESENCE OF TWO DYNAMIC SYSTEMS IN THE PERIPHERAL AUDITORY PATHWAY

10.1. Introduction; the onset-threshold method

The results presented in the preceeding chapters have led us to the conclusion that the origin of the phase effects in octave complexes is probably to be found in the hair-cell transducer system. Physiological evidence suggests that this system has a restricted dynamic range. Single-unit recordings in the auditory nerve indicate a relatively narrow dynamic range of about 40 dB (see e.g. Kiang et al., 1965). However, the dynamic characteristics determined with the aid of the pulsation-threshold method (i.e. the pulsation threshold measured as a function of the masker level) indicate a very wide dynamic range: from threshold up to 80 dB, without any indication of saturation (Houtgast, 1974; similar results have been found in our own investigations).

One reason for this might be that in the pulsation-threshold method the test tone at the threshold has the same "internal strength" as the masker.* Saturation for the

*See Houtgast (1974b), who states that the test tone is perceived as continuous when the corresponding neural responses to the masker and the test tone are continuous.

masker would then coincide with saturation for the test-tone and would not be measurable by this method. Forward-masking measurements, in which two-tone suppression is involved as well*, might offer a way out of this difficulty. In these experiments the levels of the test tone at threshold are much lower than those found with the pulsation-threshold method. Since the dynamic characteristics for masker and test-tone differ, saturation may be measurable in forward masking experiments.

Another possible explanation is that the system as studied with the aid of the pulsation-threshold method consists of two (or more) sub-systems which are excited successively. When the stimulus level is increased saturation might occur in the most sensitive system at a stimulus level for which the second system starts to be excited. It might be possible to separate these sub-systems on the basis of a difference in adaptive properties, i.e. if the first sub-system decays more slowly than the second one. This could be checked by measurements involving a variable time interval Δ between masker and test tone.

We, therefore, decided to carry out forward-masking experiments at different values of Δ, in order to determine whether saturation can be found under these conditions or whether different dynamic systems with different adaptive properties are involved. The results obtained in this way will be projected back on the pulsation-threshold results.

We have also developed a new method using the same stimulus configuration as in the pulsation-threshold method, but with an adjustable time interval between masker and test-tone pulse. We call this the "onset-threshold method". It has the same sensitivity as the pulsation-threshold method,

*See Houtgast (1973) and Shannon (1976). Shannon calls the decrease in forward masking due to the addition of a second masker tone "unmasking".

but differs in the detection criterion used. This detection criterion is the same for all values of Δ (in practice, we took Δ=50 ms) and is related to the onsets of the test tone. At high test-tone levels these onsets are clearly audible. As the level is lowered, the onsets become weaker and finally inaudible. The highest test-tone level at which the onsets can no longer be heard is defined as the onset threshold. The time interval Δ is introduced symmetrically on both sides of the test-tone pulse, for reasons explained below. For Δ=0 ms the onset criterion and the continuity criterion as used in the pulsation-threshold method give the same results to within a few decibels.

Before going on to describe our experiments, we will describe a method which may be used for interconversion of the results obtained with the pulsation-threshold method, the onset-threshold method and forward masking, so as to permit comparison of the experimental results.

10.2. A method for interconversion of the pulsation-threshold, onset-threshold and forward masking data

The experimental results of Houtgast (1973) and Shannon (1976) have demonstrated that all non-simultaneous masking procedures are basically measuring the same internal representation. We will now explain this assertion, with reference to a hypothetical system with a large dynamic range of 100 dB. Fig. 10.1a shows schematically the decay of the response after the termination of a stimulus (frequency f_s) for a certain test-tone frequency (f_t). The output in dB with respect to an arbitrary reference level is plotted against time T, and the input levels of the stimulus (L_s) are indicated against various decay curves*. The output

*In order to avoid confusion with the experimental results, in the discussion of the hypothetical systems both here and with reference to Fig. 10.8., the symbols L_s, f_s and T are used to indicate the stimulus level and frequency, and the decay time, respectively.

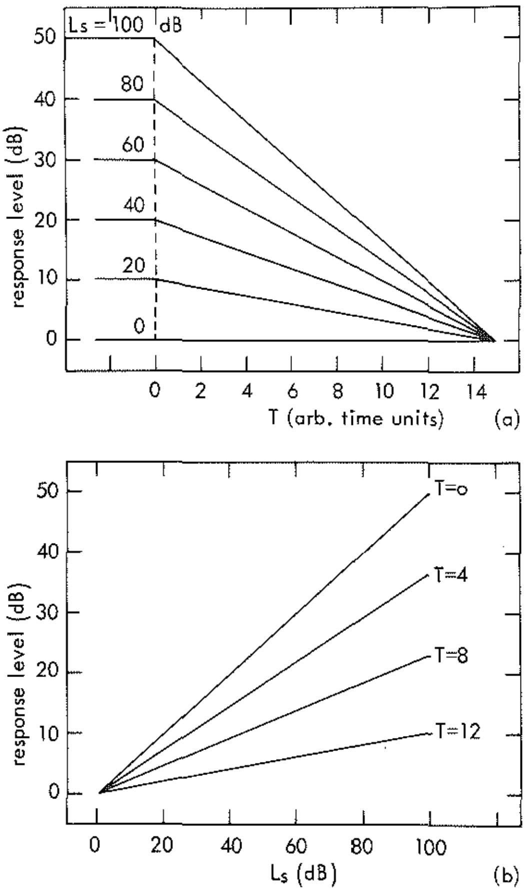


Fig. 10.1.

Decay curves and dynamic characteristics for a hypothetical system with a wide range (100 dB). L_s is the input level (in dB with respect to an arbitrary reference level), and T the time after the termination of the stimulus in arbitrary units. The termination of the stimulus is indicated by the vertical broken line.
(a) Response level and decay of the response after the termination of the stimulus.
(b) Dynamic characteristics for different values of T, derived from (a). The characteristic for T=o reflects the situation during the presence of the stimulus.

at T=o reflects the response during the presence of the stimulus.

In the pulsation-threshold method the threshold is assumed to be reached when the test-tone input level L_t is adjusted until the response to the test tone equals that to the stimulus in the frequency region f_t at a certain value of L_s. Inspection of Fig. 10.1.a. indicates that for L_s=80 dB the pulsation threshold is reached when L_t is adjusted until the response to the test tone equals 40 dB. When $f_t=f_s$ the pulsation-threshold method yields the trivial result $L_t=L_s$.

In practice there will be a difference of a few decibels due to the just noticable difference in level. In any case, the pulsation-threshold method can be used to determine the dynamic characteristic of the system, the response being expressed in terms of the input level of the test tone.

In a forward-masking set-up, with the above f_s/f_t combination, L_s=80 dB and e.g. T=12, the system gives a response level of 8 dB*. In order to compare this method with the pulsation-threshold method we are not so much interested in the 8 dB response at f_t but in the input level of a stimulus of frequency f_t needed to give a 40 dB response at T=0. This can be found in the forward-masking set-up by determining the level of a stimulus of frequency f_t that evokes the 8 dB response in question at f_t and T=12. The output for L_s=80 dB and T=12 is then expressed at the equivalent level (L_{eq}) of this stimulus. In this procedure it is assumed that a given response level always decays in the same manner, independently of the way in which it is evoked. With the system indicated in Fig. 10.1, for each f_s/f_t combination the procedure gives the same characteristic as found by the pulsation-threshold method for all values of T. The procedure is trivial for $f_s=f_t$, yielding $L_{eq}=L_s$ at each value of T, which is - within a few decibels - also the result found with the pulsation-threshold method. Thus, for the system assumed in Fig. 10.1. the correct dynamic characteristic can be measured both by forward masking and by the pulsation-threshold method. The transformation procedure just described is applicable to the onset method as well.

*Output variations of 10 dB at T=0 fall to 2 dB at T=12 in our example, i.e. the sensitivity of the method decreases with increasing T, as indicated schematically in Fig. 10.1.b. Moreover, even at T=0 forward masking can have a lower sensitivity than the pulsation-threshold method because of the different criteria used in the two methods.

10.3. Experimental method

The experiments were carried out with pure-tone maskers, either of frequency $f_m=f_1$ and level L_1 or of frequency $f_m=2f_1=f_2$ and level L_2. The test tone had the frequency $f_t=2f_1$. These stimuli are the separate components of octave complexes. In a later stage, they were presented simultaneously.

As explained in section 10.2, the difference in sensitivity of the three methods can be compensated for by expressing the pulsation threshold, the onset threshold and the forward-masked threshold of the test tone in terms of the equivalent level L_2 (L_{eq}) of a stimulus tone of frequency f_2 needed to obtain the same threshold. The procedure is illustrated with idealized diagrams in Fig. 10.2. For the pulsation-threshold method the transformation procedure implies a shift of only 1 to 2 dB.

The stimulus conditions for the pulsation-threshold and

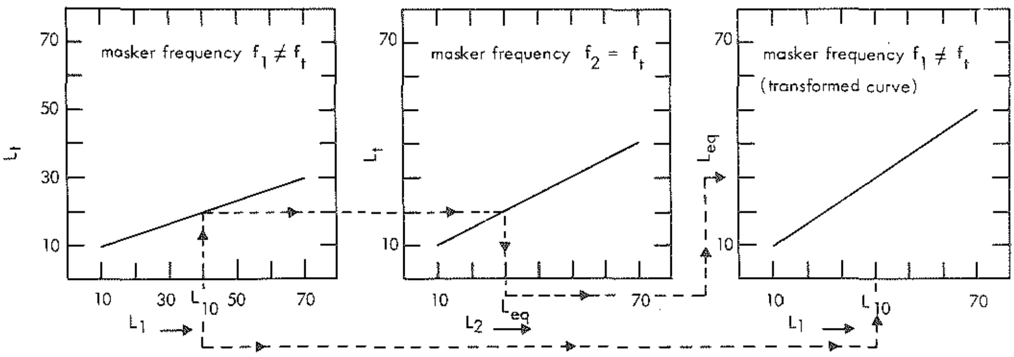


Fig. 10.2.

Transformation procedure for expressing the masked threshold L_t of a test tone of frequency f_t for a pure-tone masker of frequency $f_m \neq f_t$ in terms of the equivalent level L_{eq} of a pure-tone masker of frequency $f_m=f_t$ needed to give the same masked threshold. The procedure is illustrated for a level L_{10} of the masker tone of frequency $f_m \neq f_t$. The transformation from the value of L_t to L_{eq} is indicated by the broken lines.

forward-masking methods have been described in Chapter 3. As we mentioned in section 10.2, the stimulus conditions in the onset method are identical with those in the pulsation method except for a gap of width Δ on each side of the test-tone pulses. The shape of the gaps was the same in all three methods. The forward-masking experiments were performed by a 2-AFC procedure. Two observers participated in the experiments.

10.4. Results

10.4.1. *Comparison of pulsation-threshold method with forward masking at Δ=0 ms*

Using pure-tone maskers of frequency 760 and 1520 Hz, we measured the forward-masked threshold at Δ=0 ms and the pulsation threshold as functions of L_1 and L_2 respectively, for f_t=1520 Hz.

The results for f_m=760 Hz were transformed by expressing L_t in terms of the equivalent level L_2 (f_m=1520 Hz) needed to reach the same value of L_t, according to the procedure illustrated in Fig. 10.2. The transformed dynamic characteristics for f_m=760 Hz obtained by the two methods are shown in Fig. 10.3.a. for observer PL and in Fig. 10.3.b. for observer HK. We may note that observer HK performed the pulsation-threshold experiment with stimulus and test-tone durations of 125 ms because he was not able to perform the experiment for a duration of 300 ms. The experimental points indicated by circles correspond to the pulsation threshold, the dots to the forward-masking results.

The forward-masking experiment for the 1520 Hz masker was very difficult to perform, owing to the fact that the test tone followed the masker pulse immediately. In order to facilitate the experiment, a frequency difference of 10 Hz was introduced between masker and test-tone. Extensive measurement of the psychometric curve of the threshold with

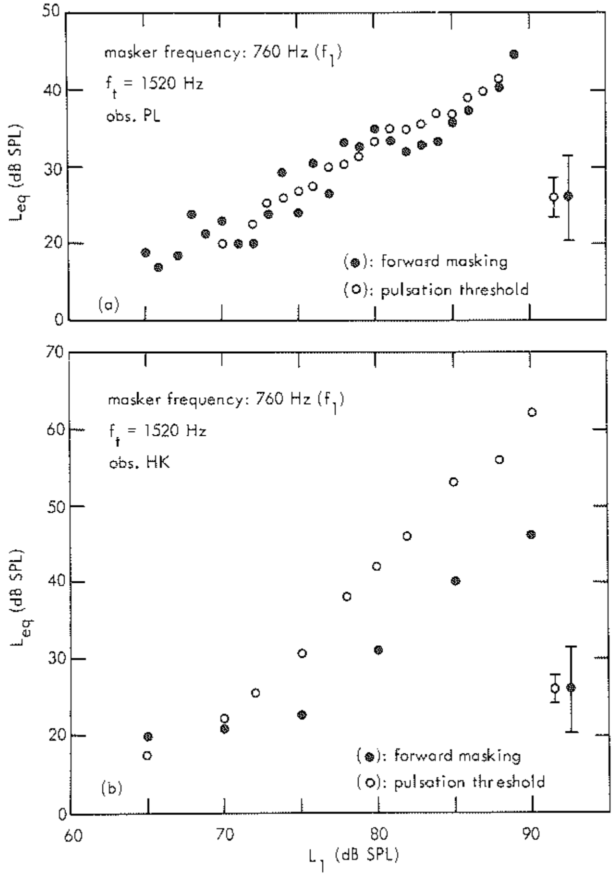


Fig. 10.3.

Forward-masked threshold (Δ=0 ms) and pulsation threshold as functions of L_m.
(a) for observer PL
(b) for observer HK
Masker frequency 760 Hz. Test-tone frequency 1520 Hz. The thresholds are given directly in L_{eq}, obtained by the transformation procedure explained in section 10.2. For the pulsation threshold the transformation only implies a shift of the data over 1 to 2 decibels.

the 2-AFC procedure revealed that two different detection criteria could be used, one corresponding to a lengthening of the masker pulse and a second one which gave rise to a lower masked threshold (6 dB down at L_2=50 dB SPL). The observers used the lower detection criterion since it is reasonable to assume that this represents the "true" response level.

Comparison of Fig. 10.3.a and 10.3.b shows that the transformed dynamic characteristics measured by the two methods are very similar for observer PL but differ somewhat for observer HK. This difference might be due to the higher repetition frequency of the stimuli in the pulsation-threshold method. This assumption is supported by the fact that the pulsation-threshold characteristic for observer HK in Fig. 10.3.b. is about 10 dB higher than that for observer PL in Fig. 10.3.a, for which the presentation was 300 ms. In order to study the influence of the duration, a pulsation-threshold experiment was performed with an intermediate presentation time of 200 ms. The average threshold for observer HK was then about 10 dB lower than the transformed forward-masking characteristic. Thus, although the quantitative agreement between the two types of dynamic characteristics is not bad, the influence of possible differences in the conditions used in the pulsation-threshold method have to be taken into account.

We may conclude that the dynamic characteristics measured by the pulsation-threshold and forward-masking methods are more or less the same after a correction has been made for the difference in sensitivity as described in section 10.2. Furthermore, the characteristics at $\Delta=0$ ms show no sign of a plateau. This indicates that the first hypothesis mentioned in section 10.1 has to be rejected: even when the "internal" level of the test tone at threshold is much lower than that of the masker, saturation is absent.

We have not considered the onset-threshold method so far because for $\Delta=0$ ms the results obtained with it do not deviate by more than a few decibels from the pulsation-threshold results.

10.4.2. *Forward masking and onset threshold for $\Delta=50$ ms*

Dynamic characteristics were now measured for $\Delta=50$ ms by forward masking and by the onset method. The frequencies f_m of the pure-tone maskers were 1520 Hz and 760 Hz, and the test-tone frequency was 1520 Hz. The results obtained with

Fig. 10.4.

Forward-masked threshold as a function of L_m. Masker frequency: 1520 hZ; f_t=1520 Hz, Δ=50 ms. Obs. PL. The experiment was performed by a 2-AFC procedure. The horizontal line indicates the absolute threshold of the short test pulse.

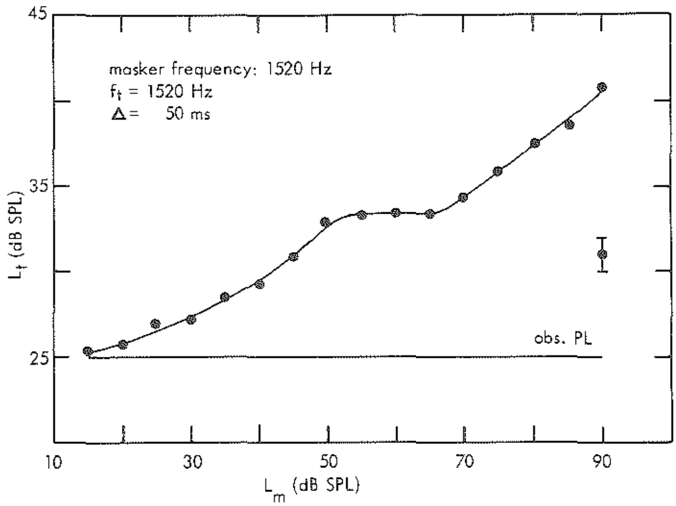
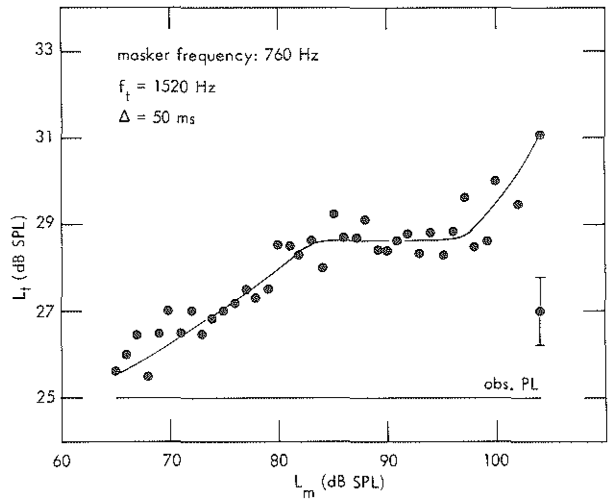


Fig. 10.5.

Forward-masked threshold as a function of L_l for f_m=760 Hz, f_t=1520 Hz and Δ=50 ms. Obs. PL. The experiment was performed by a 2-AFC procedure. The horizontal line indicates the absolute threshold of the short test pulse.

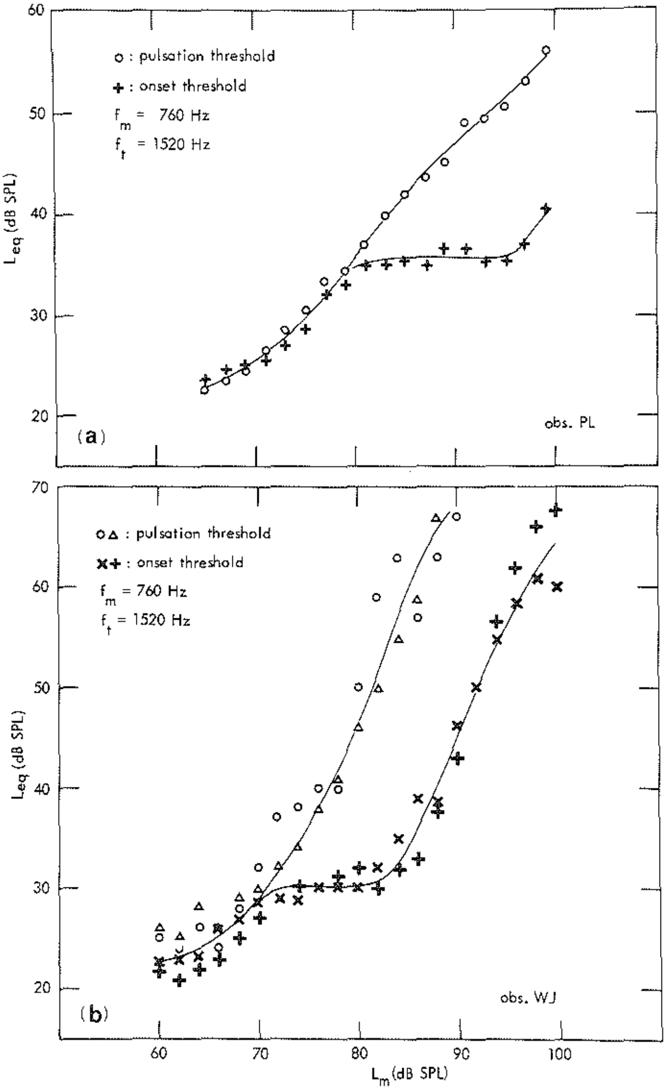


forward masking for f_m=1520 Hz are shown in Fig. 10.4 and for f_m=760 Hz in Fig. 10.5. The two curves were determined by the 2-AFC procedure.

The onset-threshold characteristics for f_m=760 Hz are shown in Fig. 10.6.a. (observer PL) and Fig. 10.6.b. (observer WJ). The corresponding pulsation-threshold characteristics are plotted alongside for the sake of comparison. The onset and pulsation thresholds are given in terms of L_{eq}

Fig. 10.6.

Onset threshold (Δ=50 ms) and pulsation threshold for a pure-tone stimulus of frequency 760 Hz with f_t=1520 Hz for two observers. The different symbols in (b) indicate the results obtained in different sessions.



directly. The transformation procedure only implied a shift of the experimental points over a few decibels*. Experimental results obtained by observer WJ for f_m=400 Hz and f_t=800 Hz.

*In the onset-threshold experiments, for $f_t=f_m$, L_m was always below the value corresponding to the start of the plateau (60 dB SPL in Fig. 10.4).

confirmed the results shown in Fig. 10.6. The two different symbols used for the onset results in Fig. 10.6.b, which represent data obtained in two different sessions, indicate that there are systematic differences of up to 6 dB in the results from session to session. However, the shape of the curves does not vary.

We now applied the method of section 10.2. to check whether the dynamic characteristics for f_m=760 Hz measured by forward masking (Fig. 10.5) and the onset-threshold method (Fig. 10.6.a) yield the same result. The curve at Δ=50 ms (Fig. 10.5), transformed with the aid of the curve for f_m=1520Hz and Δ=50 ms (Fig. 10.4), and the characteristic for Δ=0 ms (the same curve as shown in Fig. 10.3.a) are plotted together in Fig. 10.7. This result should be compared with Fig. 10.6.a. We may note that the similarity found here only holds for the symmetrical gap conditions in the onset-threshold method. When the gap between the stimulus and the subsequent test-tone pulse in the onset-threshold method was 0 ms and the other gap was 50 ms, the result was the same as in the pulsation-threshold method. On the other hand, under the reverse conditions it was very difficult to maintain the same detection criterion and a result somewhere in between the two curves of Fig. 10.6.a. was obtained. This is why we chose to use symmetrical conditions in the onset-threshold method.

It may be concluded that plateaus are present in the dynamic characteristics determined by non-simultaneous masking procedures at Δ=50 ms. With forward masking and f_t=1520 Hz the plateau is reached at L_m=60 dB SPL when f_m=1520 Hz and at 80 dB SPL when f_m=760 Hz. The height of the plateau (i.e. the corresponding value of L_t) is higher for f_m=1520 Hz (33 dB SPL) than for f_m=760 Hz (28 dB SPL). A plateau was also found with forward masking and values of Δ differing from 50 ms by four other observers.

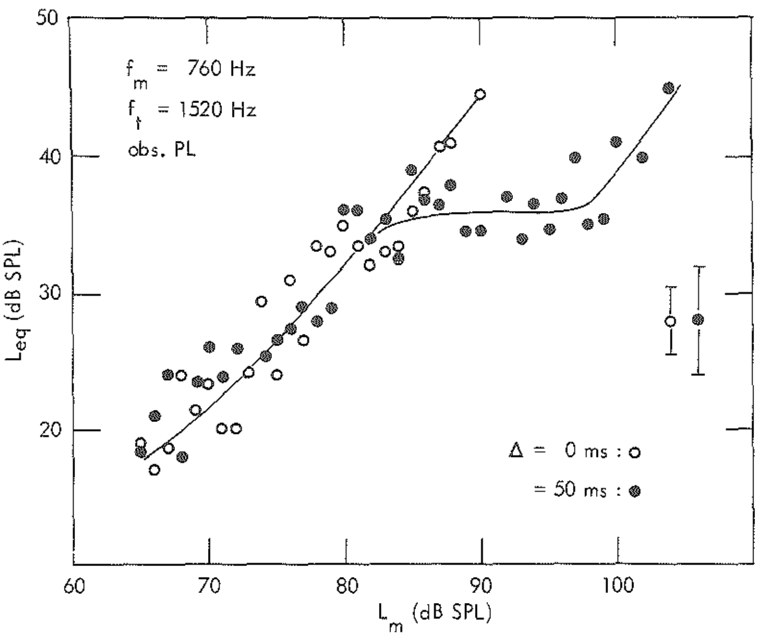


Fig. 10.7.

The same curve as in Fig. 10.5 (forward masking at Δ=50 ms), but with the threshold expressed in terms of L_{eq} of a pure tone of 1520 Hz that gives rise to the same threshold. This transformation was carried out with the help of Fig. 10.4, as described in section 10.2. The circles correspond to the transformed forward masking result at Δ=0 ms, the same curve as in Fig. 10.3.

The similarity between the transformed forward-masking characteristic for f_m=760 Hz and Δ=50 ms and the characteristic measured by the onset-threshold method implies that the latter represents a response that is characteristic for Δ=50 ms. The onset-threshold and pulsation-threshold methods have much the same sensitivity: transformation between these two only involves a shift of the data over a few decibels.

The choice of Δ=50 ms in the onset-threshold method is based on a compromise between two competing effects. On the one hand, Δ has to be as large as possible in order to obtain a distinct plateau.

On the other hand too large a value of Δ reduces the slope of the dynamic characteristic for f_t=f_m, and thus lowers the sensitivity of the method.

10.5. Discussion

The transformation procedure shown in Fig. 10.2. was applied by Houtgast (1974a) for a tone-plus-noise stimulus. His transformed forward-masking results showed the same trend as his pulsation-threshold results but were about 5 dB lower. It seems unlikely that the plateaus in the dynamic characteristics for Δ=50 ms and $f_t=f_m$ are caused by a shift of the peak in the excitation pattern. The plateaus can be found not only for test-tone frequencies of $f_t=2f_m$ but also for $f_t=f_m$+30 Hz and f_m-30 Hz.

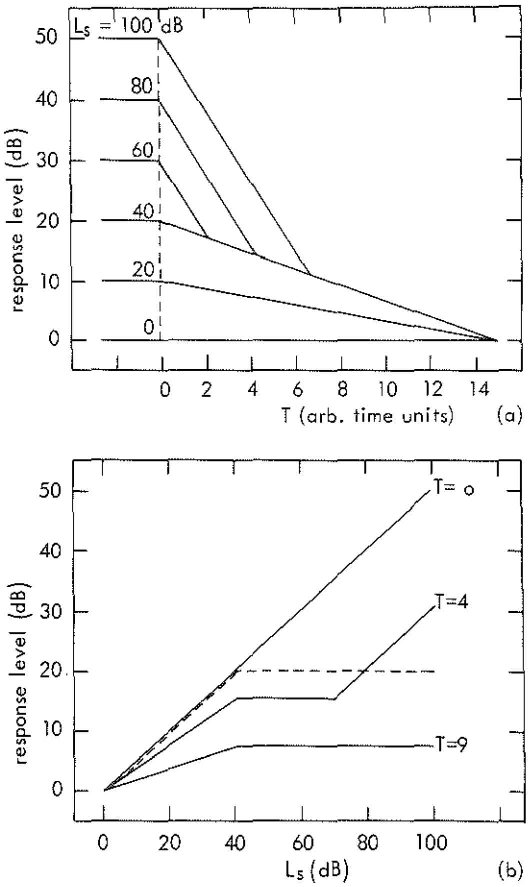
Munson and Gardner (1950) found a plateau for f_m=1000 Hz with f_t=1000 Hz and 950 Hz for Δ =100 ms in forward-masking experiments. Indications of a plateau can also be found in the data of Gardner (1947). There has been some discussion in literature as to whether these plateaus really do exist (Rawnsley et al., 1952). Referring to our remarks on the detection criterion in section 10.4.1, we may state that such a discussion should be based on data obtained by a forced-choice procedure. Moreover, L_m should be varied in small steps in these experiments because the plateau might be missed otherwise.

10.6. Interpretation of the presence of the plateaus in the dynamic characteristics for Δ=50 ms

The presence of plateaus in the dynamic characteristics for Δ=50 ms and their absence for Δ=0 ms can be interpreted as due to the involvement of two dynamic systems differing in range and adaptive properties (a fast and a slow decay), as sketched in Fig. 10.8 (cf. Fig. 10.1). The complete system consists of two parts, one with a range from threshold to 40 dB and one that takes over the excitation at L_s=40 dB. The more sensitive system has the slower decay. The dynamic characteristic of the system as a whole is represented in Fig. 10.8.b. by the straight line for T=0. The characteristic

Fig. 10.8.

Decay curves and dynamic characteristics for a hypothetical system consisting of two subsystems, one with a dynamic range from zero up to 40 dB and a second one which takes over the response at L_1=40 dB but which has a higher decay rate (see section 10.6). The symbols and scales as in Fig. 10.1.
(a) Response level during and decay of the response after the termination of the stimulus.
(b) Dynamic characteristics for some values of T derived from (a).



of the more sensitive system, which reaches saturation at L_s=40 dB, is indicated by the broken line.

It will be clear that a method which detects the response at T=0, such as the pulsation-threshold method and the forward masking method at Δ=0 ms, will reveal the dynamic characteristic of the system as a whole but will not show the presence of the more sensitive system.

However, the two systems are separated more and more by increasing the value of T (curves for T=4 and 9 in Fig.

10.8.b). The larger the value of T the more the value of L_s can be increased before the second system becomes apparent. The differences in sensitivity of the responses at the different values of T can be compensated for as described in section 10.2. On the basis of this interpretation it can be concluded from Fig. 10.4 that the more sensitive system reaches saturation at L_m=60 dB SPL when $f_t=f_m$. When $f_t=2f_m$ (Fig. 10.5) saturation in that same system is reached at L_m=80 dB SPL. In the latter case, if Δ=50 ms, L_m can be increased up to 100 dB SPL before the second system becomes apparent.

10.7. Application of the onset-threshold method to octave complexes

Let us now return to the pulsation-threshold results described in Chapter 6. As mentioned above, the pulsation method detects a response at Δ=0 ms. On the basis of the interpretation just given, this reflects the behaviour of the system in its entirety. Thus we cannot be sure which subsystem is involved for a given stimulus configuration. Moreover, since we performed many measurements as functions of L_1 and L_2, a transition can take place from one system to the other at any given level. More specifically this difficulty holds for the phase effects measured with the pulsation method. Do they originate in the more sensitive system or from both? In the latter case, which part of them originates in the more sensitive system? To answer these questions we must consider the stimulus levels and stimulus test-tone frequency combinations for which the experiments were performed. We will discuss this point with reference to Fig. 6.2. There, the stimuli are two pure tones with frequencies f_1 and $f_2=2f_1$, while the test-tone frequency f_t equals f_2. In the pulsation-threshold experiments L_2 never exceeded 50 dB SPL, so for component II the more sensitive system will not have reached saturation at this value of f_t*. However, L_1 was varied from 60 up to 87 dB SPL, so we pass the level (about 80 dB SPL) at which the more sensitive system reaches saturation for component I and the second system appears. In order to study the phase effects for the more sensitive system separately, we determined a set of curves comparable to those of Fig. 6.2. by the onset-threshold method (Δ=50 ms) for the frequency combination 760/1520 Hz with f_t=1520 Hz. In order to permit direct comparison of the results the same experiment was performed by the pulsation-threshold method. The results for L_2=35 and 45 dB SPL obtained by the two methods are shown in Fig. 10.9. Comparable results were obtained by observer WJ. The experiments were also performed for an off-ratio octave complex 800/1520 Hz with f_t=1520 Hz. These results are shown in Fig. 10.10.

Comparison of Fig. 10.9.a and b reveals that the phase effects largely originate in the more sensitive system. At L_1=80 dB SPL the second system is superimposed on the more sensitive one, so, above L_1=80 dB SPL, we cannot decide which system the phase effects determined by the pulsation-threshold method come from.

The results shown in Figs. 10.9 and 10.10 reveal a new interpretation of the phenomenon of two-tone suppression. If L_2 exceeds the value of L_{eq} (Fig. 10.10) corresponding to the plateau for component I (35 dB SPL), addition of component I and increasing its level will always cause both the pulsation threshold and the onset threshold to fall. In the results obtained by the onset method, the decrease continues until the plateau is reached. In the pulsation method, the second system becomes dominant above L_1=80 dB SPL,

*See section 10.6. We neglect the differences in frequency combination involved, as the pulsation-threshold results for these combinations are similar.

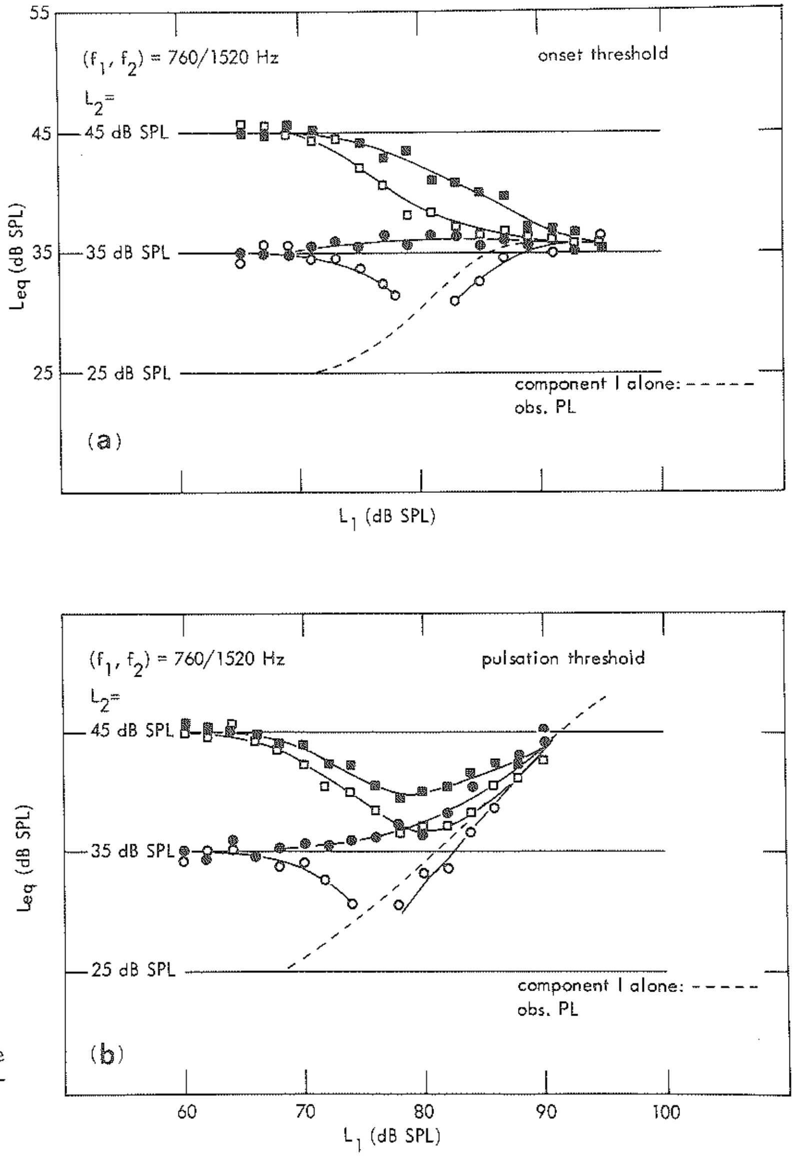


Fig. 10.9. Onset threshold (a) and pulsation threshold (b) for an on-ratio octave complex as functions of L_1 for L_2=35 and 45 dB SPL. Frequency combination 760/1520 Hz; test-tone frequency 1520 Hz. The open symbols correspond to $\phi_1=\phi_{min}$, the solid ones to $\phi_1=\phi_{max}$. The horizontal lines correspond to the pulsation and onset threshold for component II alone at the levels indicated. The broken curve corresponds to the onset threshold (a) and pulsation threshold (b) for component I alone. The reason for the absence of experimental points in the curve for L_2=35 dB SPL and $\phi_1=\phi_{min}$ in (a) and in (b) at about L_1=80 dB SPL is that no adjustments could be made, because of the low level of the test tone.

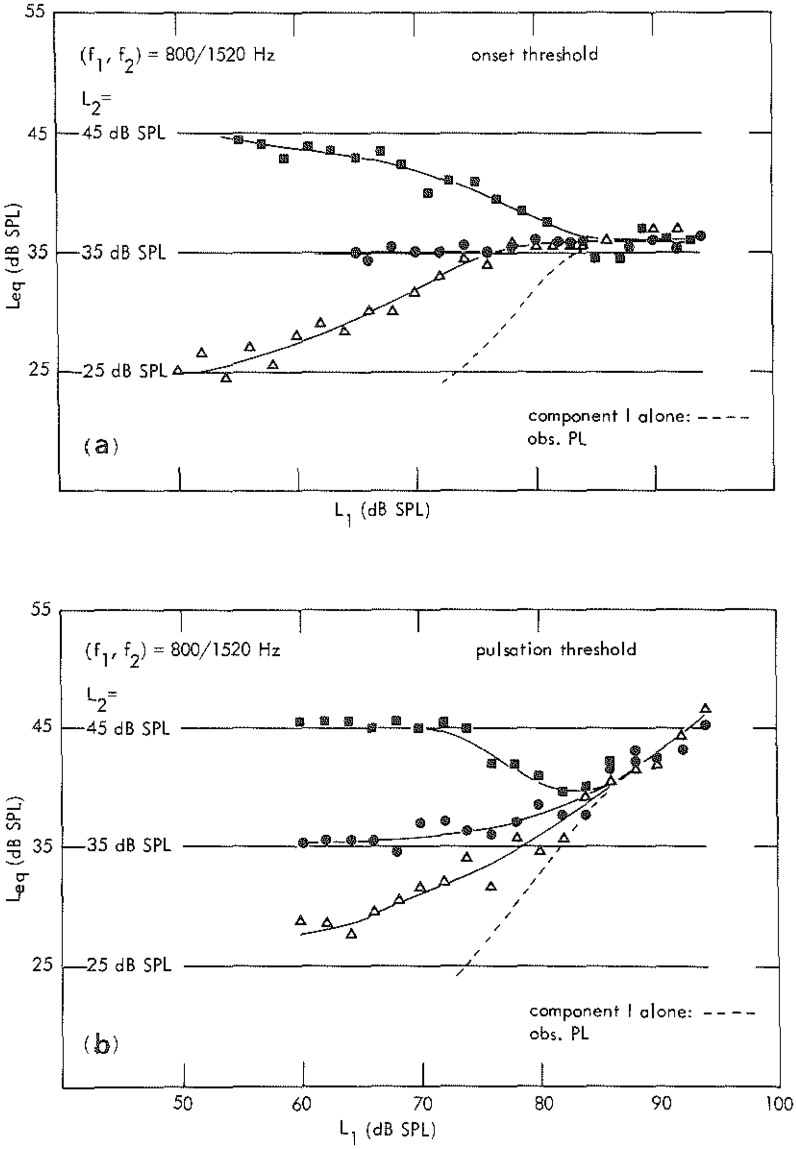


Fig. 10.10.

As Fig. 11.6., for an off-ratio octave complex 800/1520 Hz with f_t=1520 Hz.

causing the curve to rise again. If the value of L_2 is below 35 dB SPL, addition of component I will always raise the threshold. The phenomenon of two-tone suppression thus seems to be related to the plateau in the dynamic characteristic for the more sensitive system. The conclusion that both the phase effects and two-tone suppression are related to the same dynamic system supports the conclusion that these two phenomena must originate in the same stage of the auditory system (Chapter 8).

10.8. Conclusions

1) Dynamic characteristics measured by forward masking for Δ=0 ms show the same trend as dynamic characteristics measured by the pulsation-threshold method after compensation for differences in sensitivity.
2) Dynamic characteristics measured by forward masking reveal a plateau for Δ=50 ms or more.
3) The onset-threshold method can be used to measure the same response as with forward masking for Δ=50 ms, but it gives a higher sensitivity.
4) The plateau in the onset-threshold and forward-masking characteristics for Δ=50 ms may be interpreted as due to the presence of two dynamic systems differing in range and adaptive properties in the peripheral auditory system; the effects of these two systems cannot be separated in pulsation-threshold results.
5) The more sensitive of the above systems reaches saturation at L_m=60 dB SPL when f_t=f_m and at L_m=80 dB SPL when f_t=2f_m.
6) The effect of the more sensitive system can be isolated by forward-masking or onset-threshold experiments with Δ=50 ms.
7) The phase effects in octave complexes described in the Chapters 4 and 5 originate in the more sensitive system.
8) Two-tone suppression is related to the saturation level

for component I (L_{eq}=35 dB SPL) at $f_t=f_2=2f_1$. If the value of L_2 exceeds 35 dB SPL, addition of component I and increasing its level will result in a drop in the pulsation and onset thresholds.

9) The finding that both two-tone suppression and the phase effects are related to the same dynamic system supports the conclusion that they originate in the same stage of the auditory system.

CHAPTER 11

RETROSPECT; A MODEL OF THE MORE SENSITIVE OF THE TWO DYNAMIC SYSTEMS IN THE PERIPHERAL AUDITORY PATHWAY

11.1. Recapitulation of the contents of previous chapters

The two main aims of the present investigations formulated in Chapter 1 were quantification of the perceptual changes due to variation of the phase and the re-examination of the arguments for and against the aural-harmonics and waveform hypotheses. We will start this final chapter by considering to what extent these aims have been achieved.

As regards the quantification of the sensations involved in the perception of two-tone octave complexes, new elements discovered in the course of the investigations are the loudness and pitch variations of the higher-frequency component due to a change in phase, and the relationship between them. A further important element is the phenomenon of loudness reduction which is so generally reflected in our results. We have frequently stressed that the phase effects cannot be treated apart from the two-tone nonlinearity which causes loudness reduction. This means that experiments such as simultaneous masking in which two-tone nonlinearity is overlooked are unsuitable as a basis for conclusions about the possible mechanism underlying the phase effects. Another new element in our results is the relationship between the pitch effects in on-ratio and off-ratio octave complexes and between the loudness reduction (two-tone suppression) under on-and off-ratio conditions.

The aural-harmonics hypothesis has received a lot of attention in this thesis. We have shown that it is improbable that the phase effects are due to vector summation of an aural second harmonic and component II, the vector sum being "suppressed". A more likely possibility is that the phase effects are generated in the part of the auditory system where the two-tone nonlinearity originates. The conclusion of Chapter 10 that the phase effects and two-tone suppression are properties of the same dynamic system fits well within such a framework. This being so, one can debate whether one can still speak of aural harmonics as separate components or whether the more neutral term "phase effect" is more applicable. In any case, the explanation will not be complete as long as the pitch effects have no place in it.

As regards the waveform hypothesis, the relation between the phase effects and variations in the waveform of the superimposed sinusoidal components is still not very clear. Although the results for the lower-frequency combinations could be explained in terms of dispersion along the basilar membrane, the relation involved is highly frequency-dependent. It was suggested in Chapter 9 that the size of the region in which the phase effects originate might vary with frequency. However, it is by no means clear that mechanism might underly this dependence.

As regards the explanation of the pitch effects, strong indications for a pitch-place relation were found, although here again the results we obtained were frequency-dependent.

We may conclude that neither the aural-harmonics nor the waveform hypothesis can give an adequate explanation of the phase effects. The electric model presented in the next section may be considered as a condensation of our thinking about the direction in which a solution should be sought. This model describes the behaviour of the more sensitive of the dynamic subsystems mentioned in Chapter 10, for both on-ratio and off-ratio octave complexes. We shall also indicate how the model might be used for description of the pitch effects. Finally, we close this chapter with a discussion of a possible physiological basis for the model.

11.2. The model

11.2.1. *Principle*

The block diagram of the model is shown in Fig. 11.1. The model consists essentially of two combinations of a half-wave rectifier, a nonlinear amplifier and a gate, the output of one of them being delayed by a time τ and added to the other. The outputs of the gates are short pulses with a height corresponding to the positive excursions of the input signal. When the system is stimulated by a sine wave of frequency $f_p=\frac{1}{\tau}$, summation occurs since the delayed pulse train coincides with the direct one. On the other hand, when the system is stimulated with a sine wave with a frequency one octave lower, the two pulse trains (each of which has a period 2τ) are shifted in between each other. This results in a train which is similar to the above, but with half the amplitude. When a combination of these two tones with a variable phase relation is presented, the resultant amplitude varies with the phase. The peak value of the added pulse trains is taken as the magnitude of the output signal of the model.

The model may be considered as situated in a chain of similar elements, the tuning properties ($f_p=\frac{1}{\tau}$) of which change continuously as a function of distance. In the onset-threshold method the interaction in octave complexes was

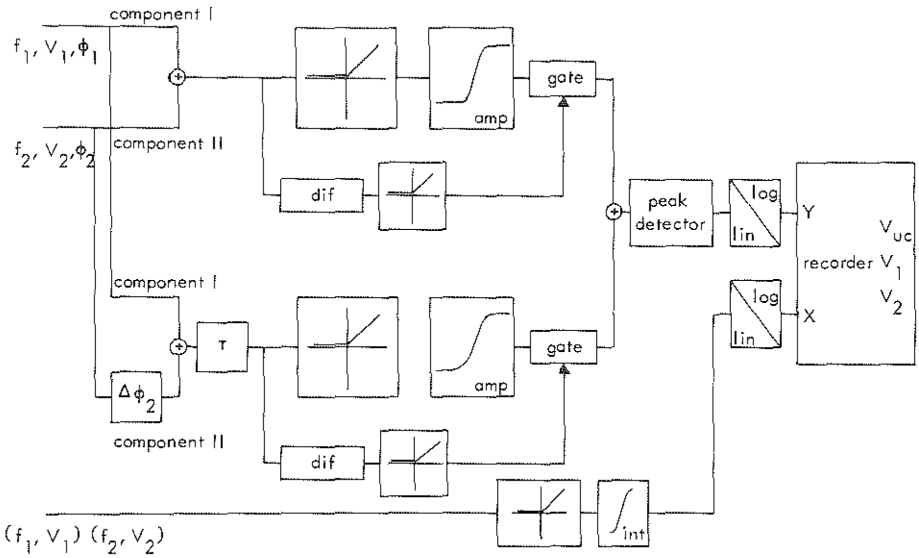


Fig. 11.1.

Block diagram of the model. For an explanation, see text.

studied for one test-tone frequency ($f_t=f_2$), which can be considered as corresponding to one element in that chain. However, this restriction cannot be maintained when we try to represent the pitch data with the aid of the model, in view of the pitch-place relation which we know to exist.

The block $\Delta\phi_2$ in Fig. 11.1. introduces a difference in phase relation in the complex input signals of the two parts of the model. This block was only used in the description of the pitch data.

11.2.2. *Design*

The model (Fig. 11.1) is composed of two parts, each consisting of a combination of a half-wave rectifier, a non-linear amplifier and a gate. The output signal of the non-linear amplifier is gated with the help of a differentiator and another half-wave rectifier. This set-up generates short pulses with a height corresponding to the positive excursions of the input signal. The nonlinear amplifiers have S-shaped characteristics, the higher and lower levels and slope of which are adjustable. The adjustable delay is obtained with an Allen Avionics delay line type DD 11 K.

The model was stimulated with "electric" octave complexes, either on-ratio or off-ratio, and with the separate components; the frequency combinations used were 500/1000 Hz and 545/1000 Hz in all experiments. The input RMS voltages of components I and II are denoted by V_1 and V_2, respectively.

As mentioned above the peak voltage of the sum of the two pulse trains is taken as the magnitude of the output signal of the model. This peak voltage is transformed into a DC voltage, logarithmically amplified and recorded as a function of the input voltage with a Houston X-Y recorder, type 2000. The output voltages are called V_{uc} for the octave complexes and V_{u1} and V_{u2} for components I and II respectively.

11.2.3. *Description of the onset-threshold data with the model*

First of all we presented input signals with frequencies of 1000 Hz and 500 Hz separately to the model, τ being 1 ms, and determined V_{u2} as a function of V_2 and V_{u1} as a function of V_1. The two curves obtained are represented in Fig. 11.2. by the dotted curves II and I respectively. With the aid of curve II we determined (for a frequency of 500 Hz) the value of V_2 needed to give an output voltage of $V_{u2}=V_{u1}$ for various values of V_1. We call this value of V_2 the equivalent value (V_{eq}) and express V_{u1} in terms of V_{eq}. This is the same transformation procedure as described in section 10.2. and illustrated in Fig. 10.2. The relation found between V_{eq} and V_1 was compared with the dynamic characteristic for

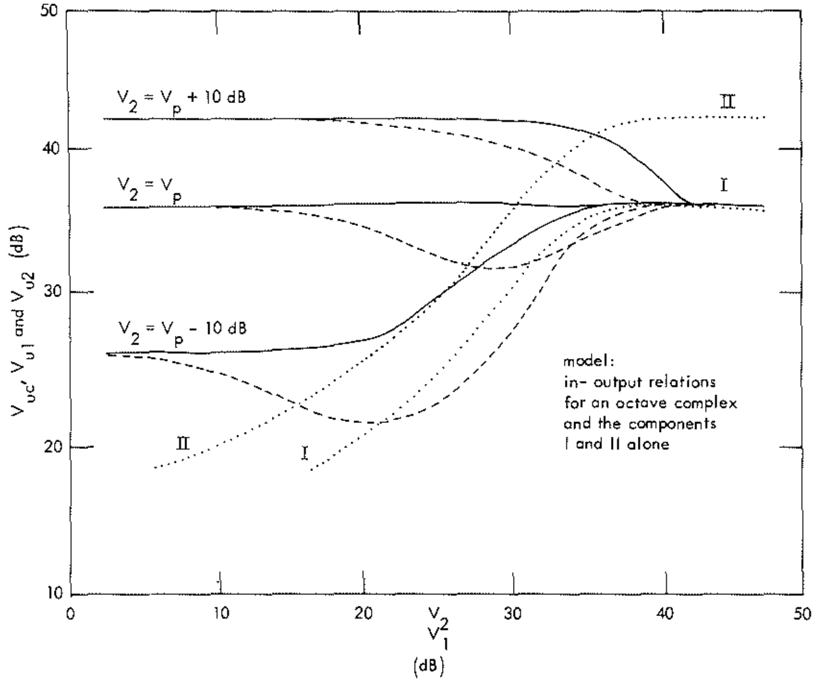


Fig. 11.2.

Input-output relations of the model for the octave complex 500/1000 Hz and the separate components. The output voltages V_{uc} for the complex and V_{u1} and V_{u2} for the components I and II respectively, expressed in dB with respect to an arbitrary reference level, are plotted against the input voltages V_1 and V_2 (also expressed in dB with respect to an arbitrary reference level). The full lines correspond to the output for the complex with $\phi_2=\phi_{2max}$, the broken lines to the output with $\phi_2=\phi_{2min}$. The dotted curve II represents the dynamic characteristic for component II alone and the curve I corresponds to the dynamic characteristic for component I alone.

component I shown in Fig. 10.9.a, and the S-shaped characteristics of the nonlinear amplifiers in the model were manipulated (both in the same way) until the two curves were identical. Now the model was ready for the presentation of octave complexes.

In the stimulation of the model with octave complexes three values of V_2 were chosen: that yielding a value of V_{u2} equal to the upper plateau in the dynamic characteristic for component I alone (which we call V_p), and also V_p+10 dB and V_p-10 dB. The stimulation with octave complexes gave rise to a phase-dependent output voltage, V_{uc} being maximum for $\phi_2=270^{\circ}$ and minimum for $\phi_2=90^{\circ}$. In the further experiments only these extreme phase conditions were considered; we will call them ϕ_{2max} and ϕ_{2min} respectively from now on. The output voltage V_{uc} for the octave complexes was determined as a function of V_1 for the three values of V_2 and two phases mentioned. The results are also plotted in Fig. 11.2. The values of V_{uc} and V_{u1} in this figure were then expressed in terms of the equivalent level (V_{eq}) of the 1000 Hz component needed to reach $V_{u2}=V_{uc}$ and $V_{u2}=V_{u1}$, respectively. The transformed data are shown in Fig. 11.3.

The model was then stimulated with the off-ratio complex 545/1000 Hz according to the same procedure as described above. The transformed results for this stimulus are shown in Fig. 11.4.*

The curves presented in Fig. 11.3 and 4 should be compared with those shown in Fig. 10.9.a. and 10.10.a. The similarity is striking. After being adjusted to give a fit for component I alone, the model describes the phase effects without the need of further assumptions. There is also a fit with respect to some details. First, the off-ratio results resemble the on-ratio results for $\phi_2=\phi_{2max}$. This resemblance was mentioned in Chapter 6. Furthermore, the output for the complex for $\phi_2=\phi_{2min}$ is lower than that for

*In Fig. 11.2. the curves for component I and component II become flatter as V_1 decreases. This implies that the phase effects for the lowest value of V_2, if expressed in terms of the equivalent level, would be very large; this is not a realistic behaviour near the auditory threshold. This effect may be a consequence of the absence of a threshold in the model. In the transformation procedure the difficulty was overcome by a vertical shift of the curve for component I alone and that for the complex at $V_2=V_p$-10 dB and $\phi_2=\phi_{2min}$.

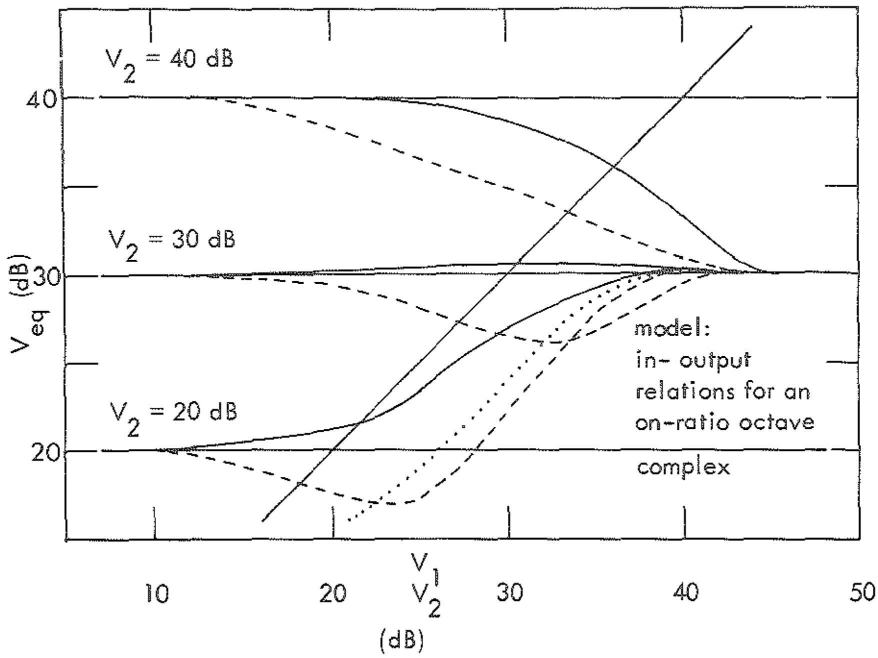


Fig. 11.3.

As Fig. 11.2, but with the output expressed in terms of V_{eq} according to the transformation procedure explained in section 10.2. The three horizontal lines correspond to the output for component II of the complex alone at the input voltages indicated. The full lines correspond to ϕ_{2max}, and the broken lines to ϕ_{2min}. The dotted curve represents the dynamic characteristic for component I alone. The oblique straight line represents the trivial dynamic characteristic for component II.

component I alone for certain values of V_1. This effect was also mentioned in Chapter 6.

The decay of the curves for the highest value of V_2 in Fig. 11.3. and 11.4. is a consequence of the elimination of summation due to the addition of component I. A similar decay for L_2=45 dB SPL in Fig. 10.9.b. and 10.10.b. was identified with the effect of two-tone suppression. With reference to the model, the two-tone suppression may, therefore, be considered as equivalent to elimination of summation. It

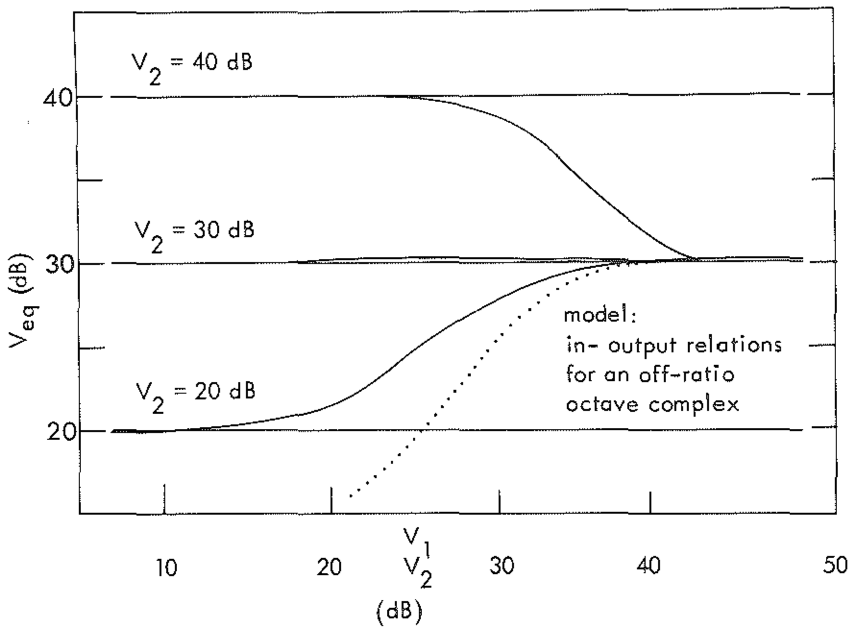


Fig. 11.4.

The transformed input-output relations for the off-ratio octave complex 545/1000 Hz. The symbols have the same significance as in Fig. 11.3.

occurs when the output of the model for component II, due to summation, exceeds the maximum output for component I (the saturation level).

The results obtained with the model also show "enhancement" for lower values of V_1 and V_2, for both on- and off-ratio stimuli, in the same manner as in our pulsation- and onset-threshold data. The model thus describes both the phase effects and the two-tone interaction effects found in our onset-threshold experiments, to a good accuracy.

11.2.4. *Suggestions for the pitch representation with the model*

In our description of the pitch data we must start from the concept of the pitch-place relation, in which a given pitch corresponds to a particular place in some internal representation of the acoustic stimulus. As mentioned above, the model is considered to be situated in a chain of similar elements, the tuning properties ($f_p=\frac{1}{\tau}$) of which change continuously with distance. Now we assume that the pitch is associated with the element in the chain at which the output signal fulfils a certain condition. As the phase shifts, different elements in the chain will be involved.

In concrete terms, let us suppose that the pitch is associated with the element in the chain with a delay τ such that the output signal of the model has a maximally regular time structure i.e. the interpeak distances in the pulse train are as equal as possible. If this condition of maximum regularity is not fulfilled at $f_p=f_2$ for an octave complex (f_1,f_2), the system scans the region around f_2 to see whether the condition is fulfilled at some other value of f_p. If this is the case, the frequency of the pure tone for which that place is maximally sensitive yields the pitch.

This principle can be illustrated by introducing $\Delta\phi_2$ (see Fig. 11.1) as the parameter. Let us take $\Delta\phi_2=60^o$. Note that $\Delta\phi_2$ only refers to component II whereas there is no phase difference in the input signals for component I alone. When we stimulate the model with a 1000 Hz component, τ must be made 1170 μs in order to get the most regular time structure. The addition of component I disturbs the regular time structure to an extent depending on the phase relation. We determined as a function of phase, with V_1-V_2 as parameter, the values of τ which give the most regular time structure. The value $\frac{f_p-f_2}{f_2}$, where $f_p=\frac{1}{\tau}$ and $f_2=\frac{1}{1170\mu s}$, are plotted against the phase in Fig. 11.5. The figure shows that the approach described above does yield a phase-dependent "pitch" with a magnitude depending on V_1-V_2. However, the figure shows a symmetrical relation between the frequency f_p for

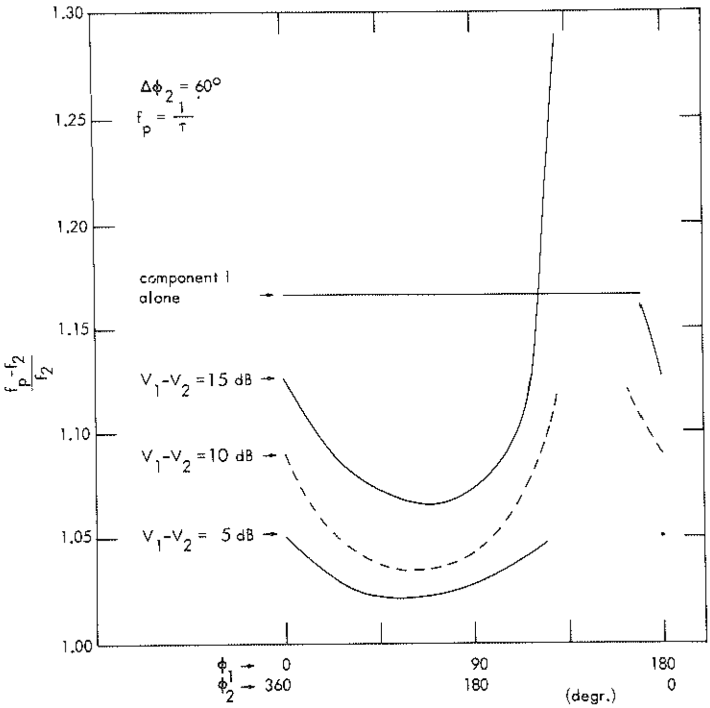


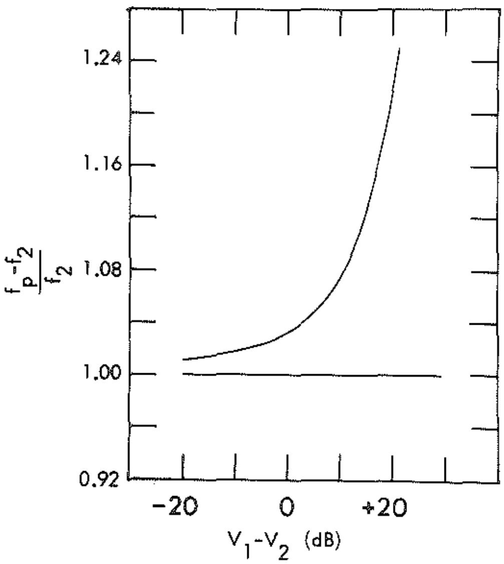
Fig. 11.5.

Relative shift, with respect to f_2, of the frequency f_p ($f_p=\frac{1}{\tau}$) at which the time structure of the output signal of the model for the octave complex 500/1000 Hz is most regular, as a function of ϕ_2 for various values of V_1-V_2.

which the time structure of the output signal is most regular and ϕ_2, in contrast to all pitch functions shown in Chapter 4. Only the part from $\phi_2=270^\circ$ to 90°, i.e. from the phase corresponding to maximum output voltage to that for the minimum output voltage, fits the experimental data.

The horizontal line in Fig. 11.5. represents the "pitch" of component I alone. This would imply that some kind of second harmonic is present in the output signal for component I alone in the form of a "pitch" which is 15% higher than that corresponding to a pure tone of frequency f_2.

Fig. 11.6.
Relative shift in f_p required to obtain the most regular output signal of the model, for the off-ratio octave complex 545/1000 Hz, as a function of V_1.



This result is in agreement with the observations of some of our subjects (see Chapter 8).

The procedure just described was also applied for the off-ratio octave complex 545/1000 Hz. We determined the frequency $f_p=\frac{1}{\tau}$ for which the time structure of the output signal was most regular by inspecting three successive peaks in the trace of the output signal on the screen of an oscilloscope and adjusting τ until these successive peaks were equidistant. The procedure was performed for various values of V_1-V_2. The resulting relative shift of f_p is plotted in Fig. 11.6. The curve obtained bears a close resemblance to Fig. 4.4.

We may conclude that the model offers possibilities for the description of the pitch effects with the aid of the concept that pitch corresponds to the most regular time structure of the output signal. However, it will be clear that more detailed assumptions (which must be based on further research) are needed to make the description complete.

11.2.5. *Possible physiological basis for the model*

A physiological basis for our model may be found in the outer hair-cell system. In this system several hair-cells are connected with one spiral fibre and arranged in a regular way. The detailed anatomy has been discussed by Spoendlin (1970). Davis (1961) was the first to suggest that spatial summation might take place in the outer hair-cell system. The basic idea of our model can be recognized in a remark made by Johnstone and Taylor (1970), who wrote, "the branching point of the nerve would act as a summing junction with the various lengths of dendrites supplying different hair cells acting as delay lines".

The application of the model to the outer hair-cell system is attended by two main difficulties. The first is the question of whether spatial summation indeed occurs. This effect requires the presence of graded potentials which propagate electronically before they are added (Davis, 1961). Flock (1973) studied recordings from the terminal branches of the sensory neurons of the burbot (lota lota). Excitatory post-synaptic potentials (EPSP's) were recorded from the non-myelinated portions. When the organ was stimulated mechanically these EPSP's were locked to one phase of the stimulus, the amplitudes being determined by the stimulus intensity. Similar observations were made by Furukawa and Ishii in the sacculus of the goldfish. However, the question as to whether these EPSP's propagate slowly enough and remain narrow enough during their course to make summation possible still remains open.

The second difficulty is the requirement that the tuning of the summation units must vary continuously along the basilar membrane. Since there is no evidence for variation of the diameter of the dendrites along the membrane, another physiological basis for this effect must be sought.

Summing up, we may state that there is reasonable agreement between the behaviour of the model and our experimental findings. However, a good deal more physiological investigation is needed to give it a firm experimental basis. More specifically, we are waiting for further results of investigations of the outer hair-cell system.

SUMMARY

Two pure tones presented simultaneously to the ear can give rise to audible beats when their frequency ratio differs slightly from 1:2. What an observer actually hears varies continuously during one period of the beats; but these various sensations can be "frozen" and studied separately by taking two pure tones with a frequency ratio of exactly 1:2 and adjusting the phase relation between the two. The nature of these sensations depends strongly on the levels of the tones. Variations in the pitch and the loudness of the higher-frequency tone can be heard when the level of the higher-frequency tone is below 40 dB SPL and that of the lower-frequency tone exceeds that of the higher-frequency one by 30 to 40 dB. These effects of phase on pitch and loudness were quantified in the present investigations.

The pitch variation as a function of the phase of the lower-frequency tone shows a sawtooth form: with increasing phase the pitch rises gradually to a maximum value (which may be up to 20% more than the initial value), drops relatively suddenly and then increases again. The magnitude of this effect is largest for the lowest frequency combinations and increases when the level of the lower-frequency tone increases with respect to that of the higher-frequency tone.

The loudness of the higher-frequency tone varies as a function of the phase between a maximum which is equal to or somewhat lower than that of this tone alone, and a minimum corresponding to the absolute threshold of the matching tone. The maximum difference between this maximum and minimum is 20 dB. The effect that the loudness even at maximum audibility can still be lower than that of the higher-frequency tone alone, was identified with the phenomenon of "loudness reduction" known from the literature. This phenomenon of "loudness reduction" is due to the two-tone nonlinearity.

Further experiments showed that the pitch and loudness (loudness reduction) effects in octave complexes, apart from

the phase dependence, are general two-tone effects. Under the same level conditions the effects in octave complexes and those in two-tone stimuli with a frequency ratio different from 1:2 are qualitively and quantitatively related.

The pulsation-threshold method, recently developed by Houtgast (1972), was used to investigate the internal representation of the octave complexes in the auditory system. The results revealed effect of two-tone suppression known from the literature. It was found that suppression in the pulsation-threshold results corresponds to reduction of the loudness of the higher-frequency tone under given stimulus conditions.

The pulsation-threshold results led to the following conclusions with respect to possible explanations of the phase effects. First, the pitch effects are not simply related to a shift of a maximum in the internal representation as determined with the pulsation-threshold method. Secondly, the hypothesis that the effects of phase on pulsation threshold and loudness are caused by interference of a second-harmonic distortion component generated in the mechanical/hydraulic part of the ear and the higher-frequency tone of the complex is not supported by the data, even when suppression of the vector sum of aural harmonic and higher-frequency tone is taken into account. The phase effects seem rather to originate in the hair-cell transducer system, the same part of the auditory system at which the two-tone nonlinearity is generated.

The third conclusion concerns the possible relation between the phase effects and variations in the waveform of the superimposed sinusoidal components. With the 200/400 Hz frequency combination and to a lesser extent with the other combinations, the phase at which the pulsation threshold is maximum depends on the test-tone frequency. This dependence can be expected when dispersion (i.e. the effect that velocity depends on frequency) occurs on the basilar membrane around the place that is maximally sensitive to the higher-frequency tone. The curves found resemble qualitatively the function obtained by determining, as a function of place, on the basilar membrane and hence of frequency (frequency-place relation), the phase in the stimulus waveform needed to give a specific waveform at that place, the same for all places. Moreover, the plot of test-tone frequency for which the pulsation threshold is maximum against phase shows a striking similarity to the pitch-phase function. This indicates that the pitch is determined by the "place" in the internal representation which originates from a specific waveform on the basilar membrane. However, the relation found is strongly frequency-dependent, and is nearly absent for the combination

760/1520 Hz.

The last part of the present thesis contains arguments in support of the statement that the pulsation-threshold results involve contributions from two different dynamic systems differing in range and temporal properties. The range of these systems depends on the combination of stimulus frequency and test-tone frequency. The more sensitive, slower system is operative from the auditory threshold up to a stimulus level of 60 dB SPL when stimulus frequency and test-tone frequency are equal, and up to a level of 80 dB SPL when the frequency of the stimulus is half that of the test tone. At these levels the more sensitive system reaches saturation and the less sensitive system appears.

A new measuring technique, the "onset-threshold method", has been developed in order to investigate the more sensitive system separately. The presentation conditions in this method are very similar to those for the pulsation-threshold method except that in the onset method there is a time interval of 50 ms between masker and test tone. The method was applied to both octave complexes and two-tone stimuli in which the frequency ratio of the components is less than two, where no phase relation is involved. It was concluded that the phase effects originate in the more sensitive of the two above-mentioned dynamic systems. Suppression may be related to the finding that the saturation level of the most sensitive system is higher if the test-tone frequency equals that of the stimulus than if it is twice that of the masker.

Finally a model is introduced to describe the phase effects obtained with the onset method. This model is based on summation over a delay line. Summation occurs when the delay time equals the period of the stimulus tone. The addition of a tone with a frequency one octave lower disturbs the summation to an extent depending on the phase. The model also describes two-tone suppression as the removal of the state of summation due to the addition of the second tone.

SAMENVATTING

Wanneer men het oor gelijktijdig twee zuivere tonen aanbiedt, die bijna een octaaf in frequentie verschillen, zijn zwevingen te horen. De sensaties die in een zwevingsperiode continu worden doorlopen kunnen afzonderlijk opgewekt en bestudeerd worden wanneer men de tonen een frequentieverschil van precies een octaaf geeft en de faserelatie instelbaar maakt. Deze fase effecten in twee-toons octaafcomplexen hangen sterk af van de niveaus van elk van de tonen. Zij manifesteren zich als een variatie van de toonhoogte en de luidheid van de hoge toon als het niveau van de hoge toon kleiner dan 40 dB SPL is en het niveau van de lage toon 30 tot 40 boven dat van de hoge toon ligt. De toonhoogte- en luidheidsvariaties onder deze condities werden gekwantificeerd.

De toonhoogte verloopt als functie van de fase volgens een zaagtand: met toenemende fase van de lage toon neemt de toonhoogte langzaam toe tot een maximale waarde; dan volgt een snelle terugval waarna de toename opnieuw begint. De toon is bij maximale verschuiving altijd hoger dan de afzonderlijke toonhoogte van de hoge toon. De grootte van het effect neemt toe naarmate de lage toon sterker wordt t.o.v. de hoge en naarmate we te maken hebben met lagere frequenties van het octaafcomplex. De maximale verschuiving die werd gevonden is ca. 20%.

De luidheid varieert als functie van de fase tussen een maximale waarde, die meestal gelijk is aan of iets minder is dan de afzonderlijke luidheid van de hoge toon en een minimale waarde die overeenkomt met de absolute drempel van de vergelijkingstoon als de lage toon maar sterk genoeg is.
De maximale luidheids variatie is ca. 20 dB. Het verschijnsel dat de luidheid soms, zelfs voor de fase waarbij een maximum optreedt, geringer is dan de oorspronkelijke luidheid werd geidentificeerd als het uit de literatuur bekende fenomeen "luidheidsreductie". Dit effect is een manifestatie van de twee-toons niet-lineariteit.
Verder werd gevonden dat de toonhoogte- en luidheid-

(luidheids-reductie) effecten in octaafcomplexen niet op zichzelf staan, maar, afgezien van de fase-afhankelijkheid, een algemeen twee-toons verschijnsel zijn. Bij gelijke niveaucondities komen de effecten voor de verschillende frequentiecombinaties kwalitatief en kwantitatief met elkaar overeen.

Met behulp van een recent ontwikkelde meettechniek, de pulsatiedrempelmethode (Houtgast, 1972) werd onderzocht hoe octaafcomplexen intern in het auditieve systeem worden gerepresenteerd. In de resultaten werd veelvuldig de reeds uit de literatuur bekende twee-toons suppressie gevonden. Het bleek dat suppressie in pulsatieresultaten overeenkomt met reductie in luidheid van de hoge toon.

Uit de resultaten van de pulsatie-experimenten konden conclusies worden getrokken m.b.t. mogelijke verklaringen van de fase-effecten. Een eerste conclusie is dat de toonhoogte effecten niet op een eenvoudige wijze samenhangen met een verschuiving van een maximum in de interne representatie zoals die met de pulsatiemethode wordt gemeten. Een tweede conclusie betreft de mogelijkheid dat de fase-effecten in pulsatiedrempel en luidheid een gevolg zijn van de aanwezigheid van een tweede harmonische vervormingscomponent ("aural harmonic"). De fase-effecten zouden dan kunnen worden beschreven als een vectoroptelling van deze aural harmonic en de hoge toon van het complex. Deze verklaring wordt, zelfs al brengt men mogelijke suppressie van de vectorsom van de aural harmonic en de hoge toon in rekening, op belangrijke punten ondergraven door de pulsatiedrempelresultaten. Het is waarschijnlijker dat de fase effecten een gevolg zijn van de twee-toons niet-lineariteit en ontstaan in het haarcel transducer systeem.

Een derde conclusie betreft een mogelijke relatie van de fase-effecten met de variatie van de golfvorm van het somsignaal van de twee tonen. Bij de 200/400 Hz frequentiecombinatie en in mindere mate bij de andere combinaties bleek de fase waarbij de pulsatiedrempel maximaal is af te hangen van de frequentie van de testtoon. Een dergelijke afhankelijkheid kan men verwachten wanneer op het basilair membraan, rond de plaats die maximaal gevoelig is voor de hoge toon, dispersie optreedt. De gemeten curves komen kwalitatief overeen met de afhankelijkheid die verkregen wordt wanneer men als functie van de plaats op het membraan (dus als functie van de frequentie in het kader van de frequentie-plaatsrelatie) de fase in de stimulusgolfvorm bepaalt die nodig is om op die plaats een specifieke, voor alle plaatsen steeds dezelfde, golfvorm te krijgen. Verder bleek de frequentie van de testtoon waarbij de pulsatiedrempel maximaal is, indien uitgezet tegen de fase, een verrassende overeenkomst op te leveren met de toonhoogte-faserelatie. Dit betekent dat de toonhoogte wordt bepaald door de "plaats" in de interne representatie die afkomstig is van een bepaalde golfvorm op het basilair membraan. Het gevonden verband is echter sterk frequentie-afhankelijk: bij 760/1520 Hz is de relatie al vrijwel afwezig.

In het laatste deel van dit proefschrift worden argumenten aangevoerd voor de stelling dat in de interne representatie van de stimuli zoals die met de pulsatiedrempelmethode wordt bepaald twee dynamische systemen een rol spelen die verschillen in bereik en temporele eigenschappen. Het bereik van deze systemen hangt af van de combinatie van stimulus- en testtoonfrequentie. Het gevoeligste en langzaamste systeem is werkzaam vanaf de drempel tot een stimulusniveau van 60 dB SPL bij gelijke stimulus- en testtoonfrequentie. Als de frequentie van de stimulus de helft is van de frequentie van de testtoon is het werkzaam tot 80 dB SPL. Bij 60 en 80 dB SPL respectievelijk treedt verzadiging op. Daarboven verschijnt het minder gevoelige en snellere systeem.

Een nieuwe meetmethode werd toegepast, de "inzet-drempelmethode" waarmee het mogelijk was het gevoeligste systeem te isoleren. De presentatiecondities in deze methode lijken veel op die bij de pulsatiedrempelmethode, met het verschil dat de tijd tussen stimulustoon en testtoon 50 ms is. De methode werd toegepast op octaafcomplexen zowel als op tweetoons stimuli waarbij de frequentieverhouding van de componenten iets kleiner was dan twee. Uit de resultaten werd geconcludeerd dat de fase-effecten afkomstig zijn van het gevoeligste systeem. Verder bleek dat suppressie samenhangt met het verschil in de verzadigingsniveaus van het gevoeligste systeem voor de hoge en lage toon bij een test-toonfrequentie die gelijk is aan de frequentie van de hoge toon.

Tenslotte werd een model geintroduceerd dat bedoeld is een beschrijving te geven van de fase-effecten zoals die met de inzetdrempelmethode werden verkregen. Het model is gebaseerd op sommatie in een vertragingslijn. Sommatie treedt op als de vertragingstijd gelijk is aan de periode van de stimulustoon. Toevoeging van een toon met een octaaf lagere frequentie verstoort de sommatie in een mate die afhankelijk is van de fase. Het model beschrijft ook suppressie als de opheffing van de sommatietoestand door het toevoegen van de lage toon.

REFERENCES

Anderson, D.J., Rose, J.E., Hind, J.E. and Brugge, J.F., Temporal position of discharges in single auditory nerve fibres within the cycle of a sine-wave stimulus: frequency and intensity effects. J. Acoust. Soc. Amer. 49 (1971), 1131-1139.

Békésy, G. von, Ueber die nichtlinearen Verzerrungen des Ohres. Ann. Phys. 20 (1934), 809-827.

Békésy, G. von., The vibration of the cochlear partition in anatomical preparations and in models of the inner ear. J. Acoust. Soc. Amer. 21 (1949), 233-245.

Békésy, G. von, Sensations on the skin similar to directional hearing, beats, and harmonics of the ear. J. Acoust. Soc. Amer. 29 (1957), 489-501.

Bilsen, F.A. and Goldstein, J.L., Pitch of dichotically delayed noise and its possible spectral basis. J. Acoust. Soc. Amer. 55 (1974), 292-296.

de Boer, E. and Bouwmeester, J., Clinical psychophysics. Audiology 14 (1975), 274-299.

Bosanquet, R.H.M., On the present state of experimental acoustics, with suggestions for the arrangement of an acoustic laboratory and a sketch of research. Phil. Mag. 5th Ser. 8 (1879), 290-305.

Bosanquet, R.H.M., On the beats of consonances of the form h:1. Phil. Mag. 5th Ser. 11 (1881), 420-436 and 492-500.

van den Brink, G., The relation between binaural diplacusis for pure tones and for complex sounds under normal conditions and with induced monaural pitch shift. Acustica 32 (1975), 159-165.

Brüel, P.V., Frederiksen, E. and Rasmussen, G., Artificial ears for the calibration of earphones of the external type. Brüel and Kjaer Technical Review (1961), 3-27.

Chapin, E.K. and Firestone, F.A., The influence of phase on tone quality and loudness; the interference of subjective harmonics. J. Acoust. Soc. Amer. 5 (1934), 173-180.

Charan, K.K., Cox, J.R. and Niemoeller, A.F., Evaluation of new couplers for circumaural earphones. J. Acoust. Soc. Amer. 38 (1965), 945-955.

Chocholle, R. and Legouix, J.P., On the inadequacy of the method of beats as a measure of aural harmonics. J. Acoust. Soc. Amer. 29 (1957a), 749-750.

Chocholle, R. and Legouix, J.P., About the sensation of beats between two tones whose frequencies are nearly a simple ratio. J. Acoust. Soc. Amer. 29 (1957b), 750.

Clack, T.D., Aural harmonics: the masking of a 2000-Hz tone by a sufficient 1000-Hz fundamental. J. Acoust. Soc. Amer. 42 (1967), 751-758.

Clack, T.D., Aural harmonics: preliminary time-intensity relationships using the tone-on-tone masking technique. J. Acoust. Soc. Amer. 43 (1968), 283-288.

Clack, T.D. and Bess, F.H., Aural harmonics: the tone-on-tone masking vs. the best-beat method in normal and abnormal listeners. Acta Oto-Laryngol. 67 (1969), 399-412.

Clack, T.D., Erdreich, J. and Knighton, R.W., Aural harmonics: the monaural phase effects at 1500 Hz, 2000 Hz, and 2500 Hz observed in tone-on-tone masking when f_1=1000 Hz. J. Acoust. Soc. Amer. 52 (1972), 536-541.
Clack, T.D., Some influences of subjective tones in monaural tone-on-tone masking. J. Acoust. Soc. Amer. 57 (1975), 172-180.
Cotton, J.C., Beats and combination tones at intervals between the unison and the octave. J. Acoust. Soc. Amer. 7 (1935), 44-50.
Craig, J.H. and Jeffress, L.A., Why Helmholtz couldn't hear monaural phase effects. J. Acoust. Soc. Amer. 32 (1960), 884-885.
Craig, J.H. and Jeffress, L.A., Effect of phase on the quality of a two-component tone. J. Acoust. Soc. Amer. 34 (1962), 1752-1760.
Dallos, P. and Sweetman, R.H., Distribution patterns of cochlear harmonics. J. Acoust. Soc. Amer. 45 (1969), 37-46.
Dallos, P., "The auditory periphery". Acad. Press. New York and London, 1973, p. 406.
Davis, H., Some principles of sensory action. Physiol. Rev. 41 (1961), 391-416.
Duifhuis, H., Cochlear nonlinearity and second filter: possible mechanisms and implication. J. Acoust. Soc. Amer. 59 (1976), 408-423.
Egan, J.P. and Hake, H.W., On the masking pattern of a simple auditory stimulus. J. Acoust. Soc. Amer. 22 (1950), 622-630.
Egan, J.P. and Meyer, D.R., Changes in pitch of tones of low frequency as a function of the pattern of excitation produced by a band of noise. J. Acoust. Soc. Amer. 22 (1950), 827-833.
Egan, J.P. and Klumpp, R.G., The error due to masking in the measurement of aural harmonics by the method of best beats. J. Acoust. Soc. Amer. 23 (1951), 275-286.
Erdreich, J. and Clack, T.D., An algorithm for analysis of phase effects: application to monaural distortion product estimation. J. Acoust. Soc. Amer. 52 (1972), 1124-1126.
Flanagan, J.L., Models for approximating basilar membrane displacement - Part II. Bell. Syst. Techn. J. 41 (1962), 959-1009.
Fletcher, H., A space-time pattern theory of hearing. J. Acoust. Soc. Amer. 1 (1930), 311-343.
Flock, A., Jørgensen, M. and Russell, I., "The physiology of individual hair cells and their synapses", in: Basic mechanisms in hearing. Ed. Møller, A.; Academic Press., New York, 1973, pp. 273-302.
Fricke, J.E., Monaural phase effects in auditory signal detection. J. Acoust. Soc. Amer. 43 (1968), 439-443.
Furman, G.G. and Frishkopf, L.S., Model of neural inhibition in the mammalian cochlea. J. Acoust. Soc. Amer. 36 (1964), 2194-2201.
Furukawa, T. and Ishii, Y., Neurophysiological studies on hearing in goldfish. J. Neurophysiol. 30 (1967), 1377-1403.
Gardner, M.B., Short duration auditory fatigue as a method of classifying hearing impairment. J. Acoust. Soc. Amer. 19 (1947), 178-190.
Green, D.M., Detection of multiple component signals in noise. J. Acoust. Soc. Amer. 30 (1958), 904-911.
Guinan, J.J. and Peake, W.T., Middle ear characteristics of anaesthetized cats. J. Acoust. Soc. Amer. 41 (1967), 1237-1261.

Hall, J.L., Two-tone distortion products in a nonlinear model of the basilar membrane. J. Acoust. Soc. Amer. 56 (1974), 1818-1828.
Hall, J.L. and Schroeder, M.R., Monaural phase effects for two-tone signals. J. Acoust. Soc. Amer. 51 (1972), 1882-1884.
Hawkins, J.E. and Stevens, S.S., The masking of pure tones and of speech by white noise. J. Acoust. Soc. Amer. 22 (1950), 6-13.
Helmholtz, H.L.F. von, Ueber Combinationstöne. Ann. Phys. Chem. 99 (1856), 497-540.
Hermann, L., Zur Frage betreffend den Einfluss der Phasen auf die Klankfarbe. Ann. Phys. Chem. 58 (1896), 391-401.
Houtgast, T., Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Amer. 51 (1972), 1885-1894.
Houtgast, T., Psychophysical experiments on "tuning curves" and "two-tone inhibition". Acustica 29 (1973), 168-179.
Houtgast, T., Lateral suppression and loudness reduction of a tone in noise. Acustica 30 (1974a), 214-221.
Houtgast, T., "Lateral suppression in hearing" (thesis). Amsterdam, 1974b.
Houtsma, A.J.M. and Goldstein, J.L., The central origin of the pitch of complex tones: evidence from musical interval recognition. J. Acoust. Soc. Amer. 51 (1972), 520-529.
Johnstone, B.M., and Boyle, A.J.F., Basilar membrane vibration examined with the Mössbauer technique. Science 158 (1967), 389-390.
Johnstone, B.M., and Taylor, K., "Mechanical aspects of cochlear function", in: Frequency analysis and periodicity detection in hearing. Eds. Plomp, R. and Smoorenburg, G.F.; Sijthoff, Leiden, 1970, pp. 81-90.
Kiang, N.Y.-s, Watanabe, T., Thomas, E.C. and Clark, L.F., Discharge patterns of single fibres in the cat's auditory nerve. Cambridge, Mass., M.I.T. Press., 1965.
Kim, D.O., Molnar, C.E. and Pfeiffer, R.R., A system of nonlinear differential equations modelling basilar membrane motion. J. Acoust. Soc. Amer. 54 (1973), 1517-1529.
König, R., Ueber den Zusammenklang zweier Töne. Ann. Phys. Chem. 157 (1876), 177-237.
König, R., Ueber den Ursprung der Stösse und Stösstöne bei harmonischen Intervallen. Ann. Phys. Chem. 12 (1881), 335-349.
ter Kuile, Th. E., Einfluss der Phasen auf die Klankfarbe. Arch. ges. Physiol. 89 (1902), 333-426.
Kuriyagawa, M. and Kameoka, A., PSE tracing method for subjective harmonics measurements and monaural phase effect on timbre. J. Acoust. Soc. Amer. 39 (1966), 1263(A).
Lawrence, M. and Yantis, P.A., Thresholds of overload in normal and pathological ears. A.M.A. Arch. Otol. 63 (1956a), 67-77.
Lawrence, M. and Yantis, P.A., Onset and growth of aural harmonics in the overloaded ear. J. Acoust. Soc. Amer. 28 (1956b), 852-858.
Lottermoser, W., Bemerkungen zu den subjectiven harmonischen Teiltönen nach G. von Békésy. Akust. Z. 2 (1937), 148.
Meyer, M.F., Beats from combining a unit frequency with a mistuned multiple. Am. J. Psychol. 62 (1949), 424-430 and 592-593.

Meyer, M.F., A test of the hydraulic theory of cochlear mechanics through multiple frequency beats. Am. J. Psychol. 67 (1954), 39-55.

Meyer, M.F., Aural harmonics are fictitious. J. Acoust. Soc. Amer. 29 (1957), 749.

Moe, C.R., An experimental study of subjective tones produced within the human ear. J. Acoust. Soc. Amer. 13 (1942), 159-166.

Møller, A.R., Improved technique for detailed measurements of the middle ear impedance. J. Acoust. Soc. Amer. 32 (1960), 250-257.

Møller, A.R., Transfer function of the middle ear. J. Acoust. Soc. Amer. 35 (1963), 1526-1534.

Møller, A.R., "Fuction of the middle ear" in: Handbook of Sensory Physiology. Eds. Keidel, W.D. and Neff, W.D.; Springer-Verlag, Berlin-Heidelberg-New York, 1974, Vol. V/I, Ch. 15.

de Morgan, A., On the beats of imperfect consonances. Trans. Cambridge. Phil. Soc. 10 (1864), 129-141.

Müller, J.J., Ueber die Tonempfindungen. Ber Verhandl. Kön. Sächs. Gesellsch. Wissensch. Leipzig, Math.-Phys. Classe. 23 (1871), 115-124.

Munson, W.A. and Gardner, M.B., Loudness patterns - a new approach. J. Acoust. Soc. Amer. 22 (1950), 177-190.

Nelson, D.A. and Bilger, R.C., Pure-tone octave masking in normal-hearing listeners. J. Speech. Hear. Res. 17 (1974), 223-251.

Newman, E.B., Stevens, S.S. and Davis, H., Factors in the production of aural harmonics and combination tones. J. Acoust. Soc. Amer. 9 (1937), 107-118.

Nixon, J.C., Raiford, C.A. and Schubert, E.D., Technique for investigating monaural phase effects. J. Acoust. Soc. Amer. 48 (1970), 554-556.

Ohm, G.S., Bemerkungen über Combinationstöne und Stösse. Ann. Phys. Chem. 47 (1839), 463-466.

Opheim, O. and Flottorp, G., The aural harmonics in normal and pathological hearing. Acta Oto-Laryngol. 45 (1955), 513-531.

Pfeiffer, R.R., A model for two-tone inhibition of single cochlear-nerve fibres. J. Acoust. Soc. Amer. 48 (1970), 1373-1378.

Pfeiffer, R.R. and Molnar, C.E., Cochlear nerve fibre discharge patterns: relationship to the cochlear microphonics. Science 167 (1970), 1614-1616.

Plomp, R., Beats of mistuned consonances. J. Acoust. Soc. Amer. 42 (1967), 462-474.

Raiford, C.A. and Schubert, E.D., Recognition of phase changes in octave complexes. J. Acoust. Soc. Amer. 50 (1971), 559-567.

Rawnsley, A.I. and Harris, J.D., Studies in short-duration auditory fatigue: II. Recovery time. J. Exper. Psychol. 43 (1952), 138-142.

Rhode, W.S., Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. J. Acoust. Soc. Amer. 49 (1971), 1218-1231.

Rhode, W.S. and Robles, L., Evidence from Mössbauer experiments for non-linear vibration in the cochlea. J. Acoust. Soc. Amer. 55 (1974), 588-596.

Ritsma, R.J., Frequencies dominant in the perception of pitch of complex sounds. J. Acoust. Soc. Amer. 42 (1967), 191-198.

Robles, L., Rhode, W.S. and Geisler, C.D., Transient response of the basilar membrane in squirrel monkeys using the Mössbauer effect. J. Acoust. Soc. Amer. 59 (1976), 926-939.

Rose, J.E., Kitzes, L.M., Morton Gibson, M. and Hind, J.E., Observations on phase-sensitive neurons of anteroventral cochlear nucleus of the cat: nonlinearity of cochlear output. J. Neurophysiol. 37 (1974), 218-253.

Sachs, M.B. and Kiang, N.Y.-s., Two-tone inhibition in auditory nerve fibres. J. Acoust. Soc. Amer. 43 (1968), 1120-1128.

Scharf, B., Partial masking. Acustica 14 (1964), 16-23.

Schoonderbeek, W., "Monaurale fase effecten in twee component octaaf complexen." Internal Report Dept. Techn. Physics, Techn. Univ. Delft, 1976.

Schouten, J.F., The perception of subjective tones. Proc. Kon. Nederl. Akad. Wetensch. 41 (1938), 1086-1093.

Schouten, J.F., Ritsma, R.J. and Cordozo, B.L., Pitch of the residue. J. Acoust. Soc. Amer. 34 (1962), 1418-1424.

Schroeder, M.R., Models of hearing. Proc. I.E.E.E. 63 (1975), 1332-1350.

Schubert, E.D., On estimating aural harmonics. J. Acoust. Soc. Amer. 45 (1969), 790-791.

Shannon, R.V., Two-tone unmasking and suppression in a forward-masking situation. J. Acoust. Soc. Amer. 59 (1976), 1460-1470.

Shaw, E.A.G. and Thiessen, G.J., Acoustics of circumaural earphones. J. Acoust. Soc. Amer. 34 (1962), 1233-1246.

Shaw, E.A.G., "The external ear" in: Handbook of Sensory Physiology. Eds. Keidel, W.D. and Neff, W.D.; Springer-Verlag, Berlin-Heidelberg-New York, 1974, Vol. V/I, Ch. 14.

Smoorenburg, G.F., Audibility region of combination tones. J. Acoust. Soc. Amer. 52 (1972), 603-614.

Spoendlin, H., "Structural basis of peripheral frequency analysis", in: Frequency Analysis and Periodicity Detection in Hearing. Eds. Plomp, R. and Smoorenburg, G.F.; Sijthoff, Leiden, 1970, pp. 2-36.

Stumpf, C., Ueber die Ermittelung von Obertönen. Ann. Phys. Chem. 57 (1896), 660-681.

Stumpf, C., Beobachtungen über Kombinationstöne. Z. Psychol. 55 (1910), 1-142.

Terhardt, E., Zur Tonhöhenwahrnehmung von Klängen II. Ein Funktionschema. Acustica 26 (1972), 187-199.

Terhardt, E. and Fastl, H., Zum Einflusz von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen. Acustica 25 (1971), 53-61.

Thomson, W., On beats of imperfect harmonics. Proc. Roy. Soc. Edinb. 9 (1878), 602-612.

Thurlow, W.R., An auditory figure-ground effect. Am. J. Psychol. 70 (1957), 653-654.

Tonndorf, J., Fluid motion in cochlear models. J. Acoust. Soc. Amer. 29 (1957), 558-568.

Tonndorf, J., Harmonic distortion in cochlear models. J. Acoust. Soc. Amer. 30 (1958a), 929-937.

Tonndorf, J., Localization of aural harmonics along the basilar membrane. of guinea pigs. J. Acoust. Soc. Amer. 30 (1958b), 938-943.

Tonndorf, J., Beats in cochlear models. J. Acoust. Soc. Amer. 31 (1959), 608-619.

Trimmer, J.D. and Firestone, F.A., An investigation of subjective tones by means of the steady tone phase effect. J. Acoust. Soc. Amer. 9 (1937), 24-29.

Verschuure, J. and van Meeteren, A.A., The effect of intensity on pitch. Acustica 32 (1975), 33-44.

Walliser, K., Ueber die Abhängigkeiten der Tonhöhenempfindung von Sinustönen vom Schallpegel, von überlagertem drosselndem Störschall und von der Darbietungsdauer. Acustica 21 (1969), 211-221.

Wegel, R.L. and Lane, C.E., The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev. 23 (1924), 266-285.

Wever, E.G., Beats and related phenomena resulting from the simultaneous sounding of two tones. Psychol. Rev. 36 (1929), 402-418 and 512-523.

Wiener, F.M. and Ross, D.A., The pressure distribution in the auditory canal in a progressive sound field. J. Acoust. Soc. Amer. 18 (1946), 401-408.

Wilson, J.P. and Johnstone, J.R., Basilar membrane and middle-ear vibration in guinea pig measured by capacitive probe. J.Acoust. Soc. Amer. 57 (1975), 705-723.

Zwicker, E., Ueber die Lautheit von ungedrosselten und gedrosselten Schallen. Acustica 13 (1963), 194-211.

Zwislocki, J., Some measurements of the impedance at the eardrum. J. Acoust. Soc. Amer. 29 (1957), 349-356.

Zwislocki, J., "Analysis of some auditory characteristics" in: Handbook of Mathematical Psychology. Eds. Luce, R.D., Bush, R.R. and Galanter, E.; Wiley, New York, 1965, Vol. III, Ch. 15.

POSTSCRIPT

The investigations described in this thesis were carried out in the Department of Biological and Medical Physics of Erasmus University, Rotterdam, under the supervision of Prof. Dr. G. van den Brink, head of the department. They were started in 1970 and performed in close collaboration with the Audiophysics Group (head Dr. M. Rodenburg) of the E.N.T. Department. During the last two years the investigations were performed within the auditory research group of the Department "Physics and Technology", in which the two above mentioned groups co-operate. Regularly discussions were held with members of the Department of Biological Physics of the Technology University, Delft (Dr. F.A. Bilsen) and the "Instituut voor Zintuigfysiologie R.V.O.-T.N.O." in Soesterberg (Prof. Dr. Ir. R. Plomp).

CURRICULUM VITAE

(op verzoek van de Erasmus Universiteit Rotterdam)

De schrijver van dit proefschrift werd geboren op 23 maart 1936 te Nieuw Beijerland. Hij bezocht de R.H.B.S. te Oud Beijerland en het Chr. Lyceum te Zeist alwaar hij in 1954 het einddiploma gymnasium β behaalde. In datzelfde jaar begon hij met de studie wis- en natuurkunde aan de Rijks Universiteit te Utrecht. Het candidaatsexamen legde hij af in 1959. In 1965 deed hij doctoraalexamen experimentele natuurkunde, hoofdvak kernfysica. Het experimenteel werk voor het doctoraalexamen betrof het vervalschema van resonanties in de reactie $^{26}Mg(p,\gamma)^{27}Al$. Het werd verricht binnen de werkgroep F.O.M.-KV, hoofd Prof. Dr. P.M. Endt, onder leiding van Dr. C. van der Leun. Van 1964 tot 1965 was de schrijver als wetenschappelijk assistent in dienst van de stichting F.O.M. Na vervulling van de militaire dienstplicht kwam hij op 1 december 1966 in dienst van Z.W.O. voor het verrichten van auditief onderzoek onder leiding van Dr. G. van den Brink binnen de afdeling Animale Fysiologie (hoofd Prof. Dr. W.G. Walter) van de Rijks Universiteit te Groningen. Per 1 augustus 1967 trad hij in dienst van de toenmalige medische Faculteit Rotterdam als wetenschappelijk medewerker op de afdeling Biologische en Medische Natuurkunde, hoofd Prof. Dr. G. van den Brink.

Vanaf zijn indiensttreding is de schrijver betrokken geweest bij het natuurkunde onderwijs aan medische studenten. Vanaf 1969 tot 1975 was hij tevens als docent akoestische fonetiek verbonden aan de avondopleiding logopedie te Rotterdam, uitgaande van de Stichting Logopedische Opleidingen; vanaf 1967 opnieuw, voor het vak audiologie aan de gecombineerde opleiding akoepedie/logopedie te Utrecht.